Artificial Intelligence **Ali Raza Syed**
Lecture 7: Probabilistic Reasoning Over Time

Introduction

Previously we studied Bayesian networks as way of reasoning with uncertainty. However, we assumed a static where information doesn't change over time. Now we consider a dynamic system where new information arrives over time and look at ways to cope with reasoning under uncertainty in such systems.

At a given time t, we consider two sets of variables: X_t which are not observed (hidden variables) and E_t which are observed (evidence variables). We take a simple example where we wish to determine whether it's raining, but we can only observe whether people have umbrellas or not. X_t corresponds to R_t, a boolean hidden variable to indicate rain and E_t corresponds to U_t, an boolean observed variable to indicate whether an umbrella was sighted.

Transition and Sensor Models

Our sequence of states starts at time $t = 0$ and over time we have a sequence of variables, $X_0, X_1, X_2, \ldots$ and $E_1, E_2, E_3, \ldots$ (the evidence variables begin to arrive at time $t = 1$). Our model has two components: the transition model and the sensor model.

The transition model informs us of how the world evolves given our previous belief of the world and is spcified as $P(X_t|X_{0:t-1})$, where $X_{0:t-1} = \{X_0, \ldots, X_{t-1}\}$. In our example, this corresponds to the probability of rain today given the history of rain over previous days. However, our history can grow infinitely leading to a complex specification.

The Markov assumption allows us to simplify this: the current state depends on a finite number of previous states. A first order Markov process specifies that the current state only depends on the previous state: $P(X_t|X_{0:t-1}) = P(X_t|X_{t-1})$. A second order Markov process specifies dependence on the previous two states: $P(X_t|X_{0:t-1}) = P(X_t|X_{t-1}, X_{t-2})$. We will only make the former assumption for our examples, but note that this may not always hold in the real world. To remedy such cases, we can choose to make higher order assumptions or augment the state by including other variables.

This leaves another issue though; we now have an unbounded set of distributions: $P(X_1|X_0), P(X_2|X_1), \ldots$. We can deal with this by making an assumption of a stationary process, which specifies that although the states can change, the model remains the same over time. Thus we always have one general distribution: $P(X_t|X_{t-1})$.

The sensor model informs us of how evidence is used to update our beliefs. Again, we require a Markov assumption for the sensor model: $P(E_t|X_{0:t}, E_{0:t-1}) = P(E_t|X_t)$. The figure in slide 14 depicts the model for our rain/umbrella example.

At the outset, we require a prior probability $P(X_0)$, and we are then able to specify the joint distribution over time: $P(X_{0:t}, E_{1:t}) = P(X_0)\prod_{i=1}^{t} P(X_t|X_{t-1})P(E_t|X_t)$.

Inference

There are three types of inference tasks that we can obtain from this model.

- Filtering is used to find the belief state: $P(X_t|e_{1:t})$. An agent would use this as the basis for decision making.
- Prediction is used to find the next states:$P(X_{t_k}|e_{t:t}), k > 0$. An agent would use evalauate possible actions when evidence is unavailable.
- Smoothing is used find previous states given the evidence we currently have available: $P(X_k|e_{1:t}), 0 \leq k < t$. An agent might use this to determine past belief states and thus learn from history.

Computation

In filtering, an algorithm can maintain a current estimate use a recursive procedure to update beliefs: $P(X_{t+1}|e_{1:t+1}) = f(P(X_t|e_{1:t}, e_{t+1})$. Slides 18-20 provide an overview of how to arrive at a function $f_{1:t}$ which is the basis for the recursive procedure. It turns out that our probabilistic specification at a given time t is a function of the transition and sensor models based on the state at time $t-1$. Slides 22-26 provide an example of applying this procedure to the rain/umbrella example. Prediction uses just one part of the algoirthm for filtering. We simply update the belief states without using the evidence.

Smoothing is a kind of interpolation between states where we have available evidence. Thus the computation is split in two parts corresponding to $e_{1:k}$ and $e_{k+1:t}$. This works out to the specification $P(X_k|e_{1:t}) = \alpha f_{1:k} bk - 1 : t$. We think of this as message passing in our model where f corresponds to messages passed forward, computed using the filtering method, and b corresponds to messages passed backwards. Slides 29-34 cover the mathematical mechanics of this process and provide an example.

Other Models

We covered the most general class of models for probabilistic reasoning in a dynamic situation. There are a number of specializations for specific cases. Hidden Markov Models (HMM) use a single discrete variable to describe the state of the world. An outcome of this case is that the model computations can be simplied using matrix algorithms since the probabilities over time can be specified as transition and sensor matrices. Kalman filters are used to model continuous variables, for example motion tracking systems. The probabilistic specification uses a Gaussian prior with linear Gaussian transition and sensor models. Dynamic Bayesian Networks (DBN) are a kind of superset of Kalman filters and HMMs, etter equipped for real world systems where non-Gaussian posteriors are required. Slides 50 and 51 provide a taxonomy of the various models that fall under the general class of Bayesian networks.