Chapter 15 (AIAMA)

Probabilistic Reasoning Over Time

Sukarna Barua
Assistant Professor, CSE, BUET

Temporal Probabilistic Models

- Static world (as we considered in Bayesian network):
  - Random variables have a fixed number of states/values.
  - Values of Random variables doesn’t change over time
- Dynamic world (time is an important factor):
  - Random variables have a fixed number of states/values.
  - Values of Random variables change over time.

Temporal Probabilistic Models

- Dynamic world has a state at time
  - State is composed of a set of random variables
  - A snapshot of the state at time is a set of values of
- State is not observable
  - State is not directly observable.
  - A set of evidence variables are observable at time [evidences depends on state]
  - We may infer which state we are in from the evidence!

Temporal Probabilistic Models: Example

You want to know whether you have infection at time step .

You can measure fever, headache, stomachache at time step

-

    - Values: Yes/No [*Unobservable by agent, hidden*]

-

    - Values: Yes/No [*Observable by agent*]

Temporal Probabilistic Models

In a temporal probabilistic model, agent have:

- Environment: Partially observable
- Belief state: What is the current state as agent maintains/believes?
- Transition model: How the environment might evolve in the next time step
- Sensor model: How the observable events happen at world state?
- Decision: How the agent take action?
  - Evidence ☾ Belief state ☾ Decision

Hidden Markov Models

- A temporal probabilistic model may be called a Hidden Markov Model (HMM) when the state is represented by a discrete random variable:
- A single state variables at time t
  - Unobservable by agent [*hidden from the agent]*
- Set of evidence variables
  - Observable by agent [*known through percepts*]

Hidden Markov Models

- What happens if world state has multiple random variables?
  - Multiple random variables may be mapped to a single random variable
  - Example: <Burglary, Earthquake> makes up agent state both are Boolean.
  - Construct a single variable <BE> with four values {0,1,2,3} where
    - 0 means Burglary=T and Earthquake = T
    - 1 means Burglary=T and Earthquake = F
    - 2 means Burglary=F and Earthquake = T
    - 3 means Burglary=F and Earthquake = F

Hidden Markov Models: Example

A security guard inside a building needs to know whether it's raining outside. He can only see if someone coming in with/without an umbrella.

▪

- Values: Yes/No [Unobservable by agent]

▪

- Values: Yes/No [Observable by agent]

Transition Model

- Specifies the probability distribution of the state at time , given the previous states:

  - Assume the size of CPT when is large [*exponentially large*]
  - Problematic as number of time steps increases
  - Not practical as current state may depend only on few previous states

Markov Assumption for Transition Model

- Assumption: Current state is independent of all states given the previous number of states

)

- Markov Process: Process satisfying Markov assumption.

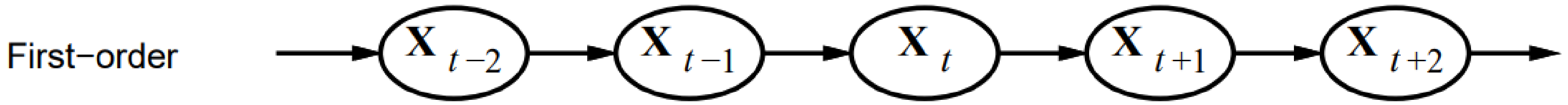
  - Also known as Markov chains.

  - After Russian mathematician Andrei Markov

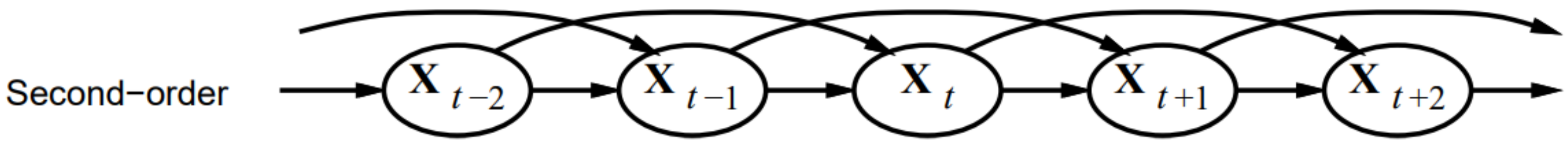
Order of Markov Process

- First Order Markov Process:
  - Current state is independent of all other states given only the previous state

  - Transition model is a conditional distribution



- For a second order Markov Process:
  - Transition model is a conditional distribution



First Order Markov Process

- Stationary process: transition model do not change over time steps

  - is same for all time steps t.

  -

  -

  [*is the probability of state transitioning from to*]

Sensor/Emission Model

- Evidence values depend on current state as well as all previous states and evidence values
- Probability distribution of events :

  - What is the probability that given all previous state and evidence values?
  - What is the size of CPT whenis large? [*exponentially large*]
  - Not practical from computational perspective

Markov Assumption for Sensor Model

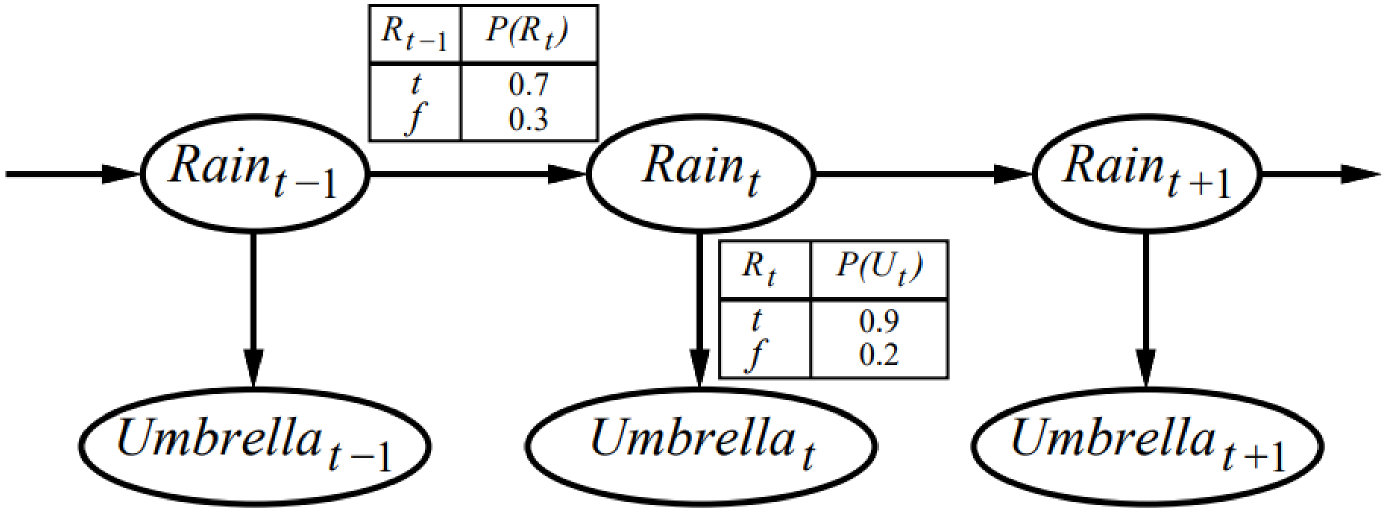
- Assumption: Evidence at time is independent of all previous states and events given the state at time (current state).

  [*evidence depend only on current state*]

  - Evidence only depend on current state and is independent of all previous states and evidences
  - [*probability of emitting output from state*]
- Also known as Observation/Emission Model

Example Markov Process

- For the umbrella example:
  - Transition model: , sensor model:



Complete/Full Joint Distribution

- We have
  - [*transition model*]
  - [*sensor model*]
- We also need
  - The prior probability distribution of states at time step
- Complete joint distribution can be computed as:

)

[Assume for notational convenience]

Complete/Full Joint Distribution

- Complete joint distribution derivation:

Is First Order Markov Process Accurate?

- Sometimes true

  - For example, in a random walk along – axis, position at time step only depends on position at time step

- Sometimes not

  - For example, in our rain example, probability of raining at time step may depend on several previous rainy days

Is First Order Markov Process Accurate?

- Sometimes not
  - For example, in our rain example, probability of raining at time step only depend on whether it rained at time step
- Solutions
  - Increase the order of the Markov process:
  - Incorporate more state variables: , etc.

Inference in First Order Markov Process

- **Filtering query**: Compute probability distribution of current state given all observations to date.
  -
  - Compute probability of raining (and not raining also!) today, given all umbrella observations taken so far
  - *Note the use capital and small letters: Capitals specify random variable and small letters specify values of random values.*
  - Required for decision making at current state

Inference in First Order Markov Process

- **Prediction query**: Compute probability distribution of a future state given all observations to date.

  - Compute probability of raining three days from now, given all umbrella observations taken so far
  - Required for decision making about future action

Inference in First Order Markov Process

- **Smoothing query**: Compute probability distribution of a past state given all observations to date.

  - Compute probability of raining last Wednesday, given all umbrella observations taken so far
  - Smoothing provides a better estimate than what was made before

Inference in First Order Markov Process

- **Most likely explanation query**: Given a sequence of observation, what is the most likely state sequence that have generated the observation sequence?

  - Umbrella was observed on first three days and absent on fourth, the most likely state sequence could be it rained first three days and did not rain on fourth.
  - Speech recognition: What is the sequence of words given a sequence of sounds?

Filtering

- Compute probability distribution of current state given observation sequence
- Agent maintains the probability distribution of current state at time step .
- As new evidence comes up, agent updates its estimation of current state probabilities

$$\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}
\end{aligned}$$

Filtering

- : a is a normalizing constant to make probabilities sum up to 1

$$
\begin{aligned}
\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)} \\
&= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}
\end{aligned}
$$

Filtering

$$\begin{aligned}\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}, \mathbf{e}_{t+1}) \quad \text{(dividing up the evidence)}\\ &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}, \mathbf{e}_{1:t})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(using Bayes' rule)}\\ &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) \quad \text{(by the sensor Markov assumption).}\end{aligned}$$

- How to calculate
  - Marginalize over

$$\begin{aligned}\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})\\ &= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \quad \text{(Markov assumption).}\end{aligned}$$

Filtering

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

$$= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \quad \text{(Markov assumption).}$$

-) comes from observation/sensor model [given]
- comes from the transition model [given]
- is the probability distribution of states at time step
  - This part is recurrence and can be computed recursively or iteratively [using dynamic programming approach]

Filtering

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t, \mathbf{e}_{1:t}) P(\mathbf{x}_t \mid \mathbf{e}_{1:t})$$

$$= \alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \sum_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{e}_{1:t}) \quad \text{(Markov assumption).}$$

- Let, [*is a vector/array of probabilities*]
  - [i] [[i] *is a single probability value*]
- Hence,

) [assume an output value]

Filtering: Forward Algorithm

-)
- is known as forward probabilities
- How to compute forward probabilities up to time step ?
  - Start from and compute [base condition]
  - Compute going forward in time up to using the recurrence
  - The algorithm is known as forward algorithm.

Filtering: Forward Algorithm

-)
- is known as forward probabilities
- How to compute compute [base condition]?

[assume is the prior probability of state]

Filering: Example

- Compute)
- Day 1:

  - is the prior probability distribution of initial state [at time

    - If both states are equally likely from START,

  - can now be calculated as:

$$
\begin{aligned}
\mathbf{P}(R_1 \mid u_1) &= \alpha\, \mathbf{P}(u_1 \mid R_1)\mathbf{P}(R_1) = \alpha \langle 0.9, 0.2 \rangle \langle 0.5, 0.5 \rangle \\
&= \alpha \langle 0.45, 0.1 \rangle \approx \langle 0.818, 0.182 \rangle \, .
\end{aligned}
$$

Filtering: Example

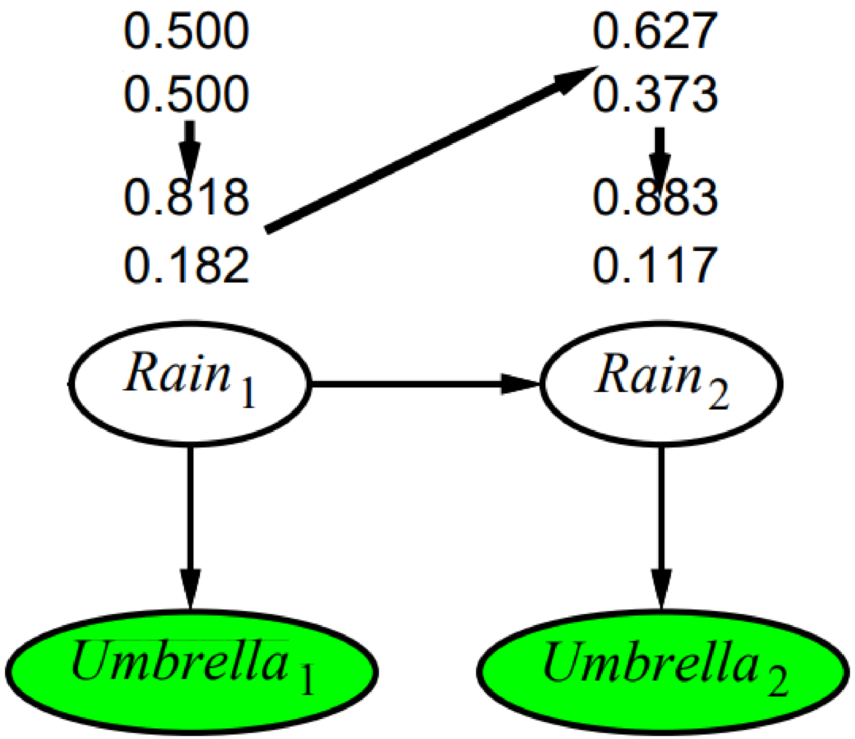
- Day 2:
  - Can be calculated as:

$$\mathbf{P}(R_2 \mid u_1) = \sum_{r_1} \mathbf{P}(R_2 \mid r_1) P(r_1 \mid u_1)$$
$$= \langle 0.7, 0.3 \rangle \times 0.818 + \langle 0.3, 0.7 \rangle \times 0.182 \approx \langle 0.627, 0.373 \rangle$$

$$\mathbf{P}(R_2 \mid u_1, u_2) = \alpha\, \mathbf{P}(u_2 \mid R_2) \mathbf{P}(R_2 \mid u_1) = \alpha \langle 0.9, 0.2 \rangle \langle 0.627, 0.373 \rangle$$
$$= \alpha \langle 0.565, 0.075 \rangle \approx \langle 0.883, 0.117 \rangle\,.$$

Filtering: Example

- Probability of rain increases at day 2 from day 1 [why?]



Prediction

- Compute probability distribution of a future state:)
- Can be computed using filtering:

  - First compute [*forward algorithm*]

  - Then compute as:

  - Similarly, compute , …,

- Recursive/dynamic programming algorithm:

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

Prediction: Don't Go Too Much Ahead

- Recursive/dynamic programming algorithm:

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

- Predicting too much ahead may be useless
  - will become fixed (stationary distribution of the Markov Process) after some time steps
  - The time taken to reach the fixed point is known as Mixing Time.

- The more uncertainty in the transition model, the shorter will be the mixing time and the more future is obscured!

Likelihood of Evidence Sequence

- What is the likelihood of evidence sequence
- Compute as
    -

- can be calculated recursively or using dynamic programming:

[*Markov assumption*]

  - can be computed recursively [using dynamic programming]
  - This is similar to the forward algorithm [described earlier]