Problem Solving with AI Techniques
Markov Models

Paul Weng

UM-SJTU Joint Institute

VE593, Fall 2018

JOINT INSTITUTE
交大密西根学院

Reasoning under Uncertainty in Dynamic Systems

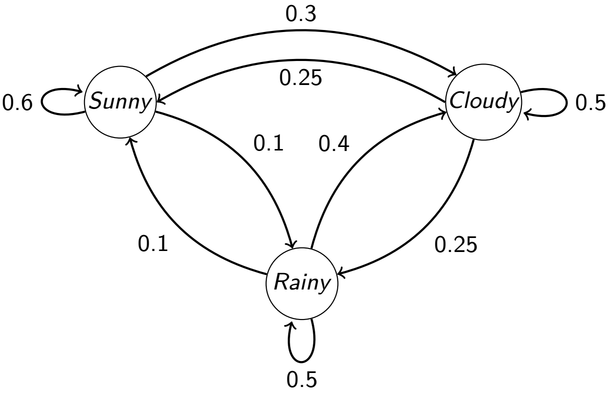
- In standard Bayes net, there's no explicit mention of time
- Insufficient to describe a dynamic system whose state may evolve (possibly stochastically)
- However, many applications require reasoning about a sequence of observations
  - Robot localization
  - Object tracking in videos
  - Speech processing and NLP
  - Medical monitoring
- A Markov model can represent how a dynamic system evolves
- Markov model = probabilistic state-based model that satisfies the Markov property:

$$\mathbb{P}(S_{t+1} \mid S_t, S_{t-1}, \ldots, S_0) = \mathbb{P}(S_{t+1} \mid S_t)$$

- Markov property: future independent of past given present

Example: Weather

$$S_0 \rightarrow S_1 \rightarrow S_2 \rightarrow \cdots \rightarrow S_t \rightarrow S_{t+1} \rightarrow \cdots$$



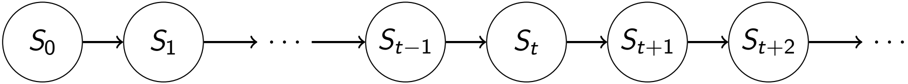
Sunny, Sunny, Cloudy, Rainy, Cloudy, Sunny...

Markov process

- Markov process is defined as a tuple
  - $\mathcal{S}$ a set of states
  - $\mathbb{P}(S_{t+1} \mid S_t)$ transition probabilities, define dynamics of the system
  - $\mathbb{P}(S_0)$ distribution over initial states
- Remarks:
  - Discrete time vs continuous time
  - Discrete vs continuous state space
  - Time-homogeneous (or stationary) vs time-inhomogeneous process
- We will focus on stationary Markov chain: time-homogeneous, discrete time, (discrete state space) Markov process

Markov Chain

- Stationary Markov Chain is defined as a tuple
  - $\mathcal{S} \subseteq \mathbb{N}$ a discrete set of states
  - $\mathbb{P}(S_{t+1} = j \mid S_t = i) = p_{ij}$ transition probabilities, with $\boldsymbol{P} = (p_{ij})$ in finite case
  - $\mathbb{P}(S_0 = i) = q_i$ distribution over initial states, with $\boldsymbol{q} = (q_i)$ in finite case



- A Markov chain defines a joint probability:

$$\begin{aligned}\mathbb{P}(S_0, S_1, \ldots, S_T) &= \mathbb{P}(S_0)\mathbb{P}(S_1 \mid S_0) \ldots \mathbb{P}(S_T \mid S_{T-1}) \\ &= \mathbb{P}(S_0) \prod_{t=1}^{T} \mathbb{P}(S_T \mid S_{T-1})\end{aligned}$$

Is the Markov Property Restrictive?

- Markov property = memoryless property
- First-order Markov property: $\mathbb{P}(S_{t+1} \mid S_t, S_{t-1}, \ldots, S_0) = \mathbb{P}(S_{t+1} \mid S_t)$
- *k*-order Markov property:
  $\mathbb{P}(S_{t+1} \mid S_t, S_{t-1}, \ldots, S_0) = \mathbb{P}(S_{t+1} \mid S_t, S_{t-1}, \ldots, S_{t-k+1})$

- Solutions if first-order Markov property is not satisfied:
  - Define new state as several consecutive states $S_t, S_{t-1}, \ldots, S_{t-k+1}$
  - Add extra information in state
- Markov model: approximation of real system

Inference Problems

With a Markov model, we can answer questions like:

- $\mathbb{P}(S_0, S_1 \ldots, S_T)$
  e.g., What is the probability of having 10 sunny days in a row?
- $\mathbb{P}(S_t \mid S_0 = s_0)$
  e.g., What is the probability of raining next week if today is sunny?
- $\mathbb{E}[T_s \mid S_0 = s_0]$ where $T_s =$ first time of reaching s
  e.g., If it rains now, what is the expected number of days before it's sunny again?
- $\mathbb{P}(S_\infty)$
  e.g., How frequently does it rain?

Basic Probabilities in Markov Chains

- State distribution at time step t: $\mathbb{P}(S_t \mid S_0 \sim q) = q\boldsymbol{P}^t$ where q distribution over states
- First passage probability of j from i at time step t

$$\begin{aligned} r_{ij}^t &= \mathbb{P}(S_t = j, \forall 1 \leq \tau < t, S_\tau \neq j \mid S_0 = i) \\ &= \sum_{s_{t-1}\neq j, s_{t-1}\neq j, \dots s_1 \neq j} \mathbb{P}(S_t = j \mid S_{t-1} = s_{t-1})\mathbb{P}(S_{t-1} = s_{t-1} \mid S_{t-2} = s_{t-2}) \\ &\qquad \dots \mathbb{P}(S_2 = s_2 \mid S_1 = s_1)\mathbb{P}(S_1 = s_1 \mid S_0 = i) \end{aligned}$$

- Probability of reaching a state j from i at some time step $t > 0$

$$f_{ij} = \sum_{t>0} r_{ij}^t$$

- Hitting time h_{ij} = expected number of steps to visit j for the first time from i

$$h_{ij} = \sum_{t>0} t r_{ij}^t$$

Classification of States

***Irreductibility*:**

- State j is reachable/accessible from state i if $\exists t > 0, (\boldsymbol{p}^t)_{ij} > 0$
- States i and j mutually reachable if i reachable from j and vice versa
- Set of states $\mathcal{C}$ is irreducible if any two states in C are mutually reachable
- Markov chain is irreducible if any two states are mutually reachable

***Transience/Recurrence*:**

- State i is persistent/recurrent if $f_{ii} = 1$; otherwise it is transient
- Set of states $\mathcal{C}$ is recurrent if all its states are recurrent
- Markov chain recurrent if all its states are recurrent

Classification of Markov Chains

- Set of states $\mathcal{C}$ is closed if $\forall i \in \mathcal{C}, \forall j \notin C, p_{ij} = 0$
- State s is absorbing if $\mathcal{C} = \{s\}$ is closed

- Decomposition Theorem:
  The state space can be uniquely partitioned in $\mathcal{S} = \mathcal{T} \cup \mathcal{C}_1 \cup \mathcal{C}_2 \cup \dots$ where $\mathcal{T}$ is the set of transient states, and $\mathcal{C}_i$ are irreducible closed sets of persistent states

Classification of Markov Chains

- Set of states $\mathcal{C}$ is closed if $\forall i \in \mathcal{C}, \forall j \notin C, p_{ij} = 0$
- State s is absorbing if $\mathcal{C} = \{s\}$ is closed

- Decomposition Theorem:
  The state space can be uniquely partitioned in $\mathcal{S} = \mathcal{T} \cup \mathcal{C}_1 \cup \mathcal{C}_2 \cup \dots$ where $\mathcal{T}$ is the set of transient states, and $\mathcal{C}_i$ are irreducible closed sets of persistent states

- Generally, we focus on irreducible recurrent Markov chains

Some Other Basic Concepts About States

- State i positive recurrent if $h_{ii} < \infty$; otherwise it is null-recurrent
- If state i reachable from itself, $d_i = \gcd(\{t \mid (\boldsymbol{p}^t)_{ii} > 0\})$ is period of i
- State i is aperiodic if $d_i = 1$
- State i is ergodic if aperiodic and positive recurrent
- Irreducible Markov chain is ergodic if all states are ergodic
- Finite Irreducible Markov chain is ergodic if it has an aperiodic state

Stationary distributions

- Stationary distribution π over states satisfies:

$$\forall j \in \mathcal{S}, \pi_j = \sum_i \pi_i p_{ij}$$

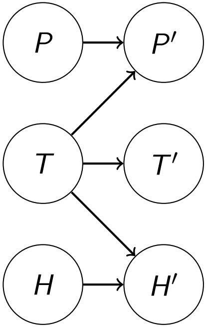
$$\pi = \pi \boldsymbol{p}$$

- It is a normalized eigenvector associated to eigenvalue 1.
- Theorem: For irreducible ergodic Markov chain, $\mathbb{P}(S_t)$ converges to a stationary distribution (called equilibrium or steady-state distribution) as $t \to \infty$

Dynamic Bayesian Network

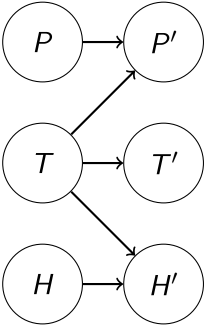
- State S is a vector of random variables $(X_1, X_2, \dots, X_n)$, its next state is $S' = (X'_1, X'_2, \dots, X'_n)$
- Transition probabilities $P(S' \mid S)$ can be represented compactly with a Bayes net by exploiting (conditional) independences
- Dynamic Bayes Net (DBN) is a Bayes net such that:
  - Nodes are $\{X_1, X_2, \dots, X_n\} \cup \{X'_1, X'_2, \dots, X'_n\}$
  - There is no edge from X'_i to X_j
  - CPT in $\{X_1, X_2, \dots, X_n\}$ = distribution over initial states
  - CPT in $\{X'_1, X'_2, \dots, X'_n\}$ = transition probabilities

Example

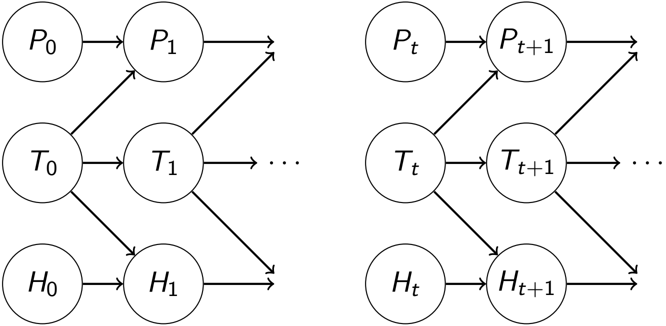


DBN

Example



DBN



BN unrolled over time

Why is DBN useful?

- By exploiting the structure of DBN:
  - Space requirement may be exponential less
  - Inference may be exponentially faster
- Inference algorithms for Bayes net can be adapted to DBN by unrolling it over time
- Structure/parameters of DBN can be learned like a Bayes net

- Issue: State is generally not fully observable!