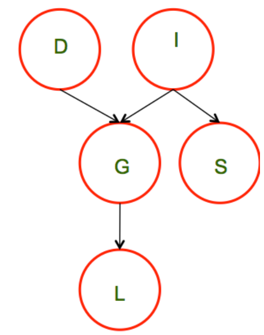
Template-Based Representations

Sargur Srihari
srihari@cedar.buffalo.edu

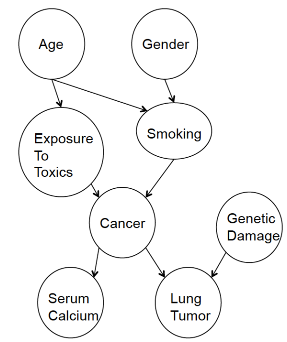
Topics

- Variable-based vs Template-based
- Temporal Models
  - Basic Assumptions
  - Dynamic Bayesian Networks
  - Hidden Markov Models
  - Linear Dynamical Systems
- Template Variables

Introduction



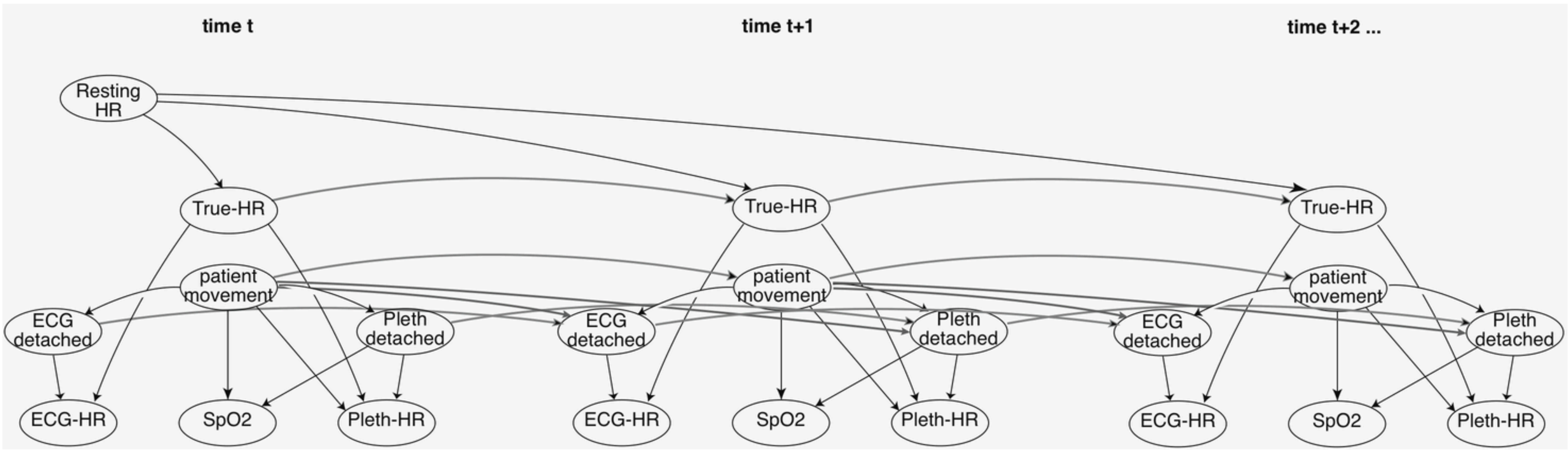
- A PGM specifies a joint distribution over a fixed set χ of variables
- A network for medical diagnosis can be applied to multiple patients, each with different symptoms and diseases
- Such networks are variable-based
- Sometimes need a much more complex space than a fixed set of variables



Temporal Setting

- Distributions of systems whose states change with time
- Ex: monitoring patient in ICU
  - Obtain sensor readings at regular intervals
    - Heart rate, blood pressure, EKG
  - Need to track over time
- Ex: Robot location tracking
  - As it moves around and gathers observations
  - Need a single model to apply to trajectories
    - Possibly of infinite length

ICU Monitoring using a DBN



- Transition model for true-HR is linear Gaussian: at time 0, μ=80 with σ= 30
- Sensor models for ECG-HR and pleth-HR: Gaussians centered at true-HR
- Sensor model for SpO2 is a Gaussian
- Transition models for ECG-detached, pleth-detached, and patient-movement assign a high probability to persisting in the current state.

Template

- Single compact model
- Provides a template for entire class of distributions
  - From the same type of trajectories
  - Temporal modeling using Dynamic Bayesian Networks
  - Or different pedigrees

Temporal Models

- State of the world evolves over time
- System State
  - Value at time t is a snapshot of the relevant attributes (hidden or observed)
  - Assignment of values to a set of random variables χ
- Use $X_i^{(t)}$ to represent instantiation of the variable X_i at time t
- X_i is no longer a variable that takes a value; rather it is a template variable

Template Variable Notation

- Template is instantiated at different points of time t
- Each $X_i^{(t)}$ takes a value in $Val\,(X_i)$
- For a set of variables $\boldsymbol{X} \subseteq \chi$

  we use $\boldsymbol{X}^{(t1:t2)}\,(t_1 < t_2)$

  to denote the set of variables

  $\{\boldsymbol{X}^{(t)}: t \in [t_1, t_2]\}$

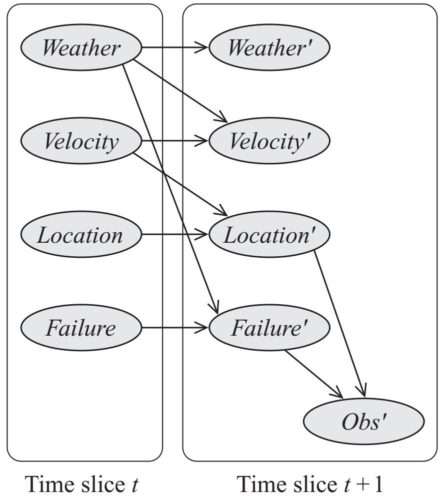
  Assignment of values to each variable $X_i^{(t)}$ for each relevant time t is a trajectory

Trajectory

- An assignment to each variable $X_i^{(t)}$
- Goal is to represent probability distributions over such trajectories
- Representing such a distribution is difficult
- Need to make some simplifying assumptions

Vehicle Localization Task

- Track current location using faulty sensors
- System state encoded using
    - X_1: *Location (car's current location)*
    - X_2: *Velocity (car's current velocity)*
    - X_3: *Weather (current weather)*
    - X_4: *Failure (failure status of sensor)*
    - X_5: *Obs (current observation)*
  - One such set of variables for every point t



(a) $\mathcal{B}_{\rightarrow}$

- Queries
    - Where is it now (t')?
    - Where is it likely to be in ten minutes?
    - Did it stop at the red light?

Basic Assumptions

- Discretize timeline into time slices
  - Time granularity
  - Set of random variables $\chi^{(0:T)}=\{ \chi^{(0)}, \chi^{(1)},..., \chi^{(T)} \}$
- Using chain rule $P(A,B,C)=P(A)P(B/A)P(C/A,B)$,

$$P(\chi^{(0:T)}) = P(\chi^{(0)})\prod_{t=0}^{T-1} P(\chi^{(t+1)} \mid \chi^{(0:t)})$$

  - Note dependence on all of time $0{:}t$
- Distribution over trajectory is the product of conditional distributions
- Need to simplify this formulation

Markovian System

- Future is conditionally independent of the past given the present
- A dynamic system over template variables χ satisfies the Markov assumption if

$$\chi^{(t+1)} \perp \chi^{(t-1)} \mid \chi^{(t)}$$

- Helps define a more compact representation of the distribution

$$P(\chi^{(0:T)}) = P(\chi^{(0)}) \prod_{t=0}^{T-1} P(\chi^{(t+1)} \mid \chi^{(t)})$$

Stationary Markovian System

- A Markovian dynamic system is stationary if $P(\chi^{(t+1)}/\chi^{(t)})$ is the same for all t.
- In this case we can represent the process using a transition model $P(\chi/\chi')$ so that for any $t \geq 0$

$$P(\chi^{(t+1)}=\xi'/\chi^{(t)}=\xi) = P(\chi'=\xi'/\chi=\xi)$$

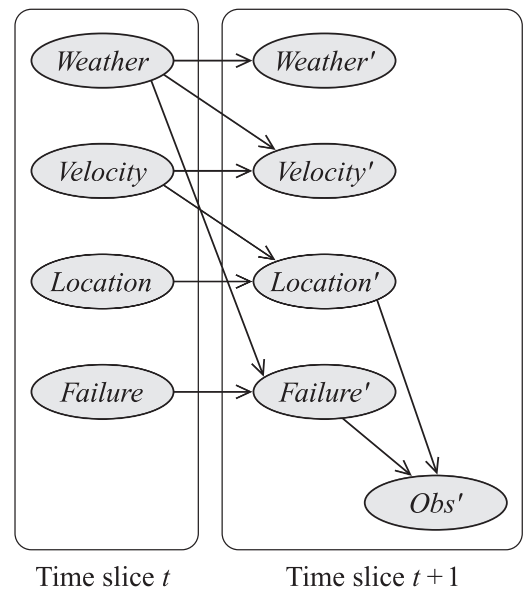
- Non-stationarity: if the conditional distribution changes with t
  - E.g., variables in biological systems
    - Change more in early years than in later years

Dynamic Bayesian Networks

- Stationarity allows representing probability distributions over infinite trajectories
- Transition model $P(\chi'|\chi)$ can be represented as a *conditional* Bayesian network

DBN for monitoring a vehicle

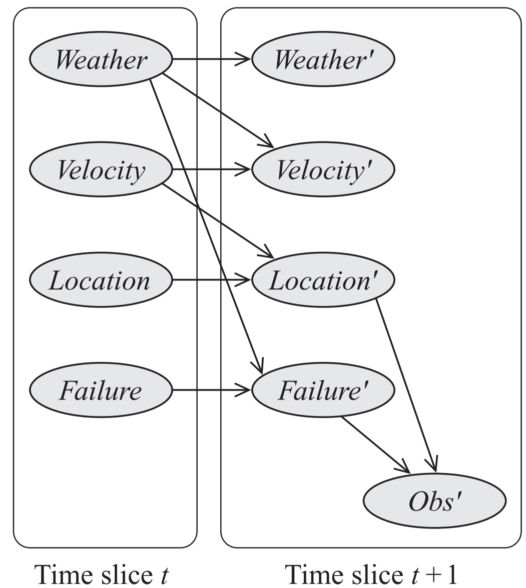
- Represents system dynamics
- X_5: *Observation* depends on car's location (and map not modeled) and error status of sensor (*failure*) (X_4)
- X_1: Bad *weather* makes sensor likely to fail (X_4)
- X_3: *Location* depends on previous position and *velocity* (X_2)



(a) $\mathcal{B}_{\rightarrow}$

Interface Variables

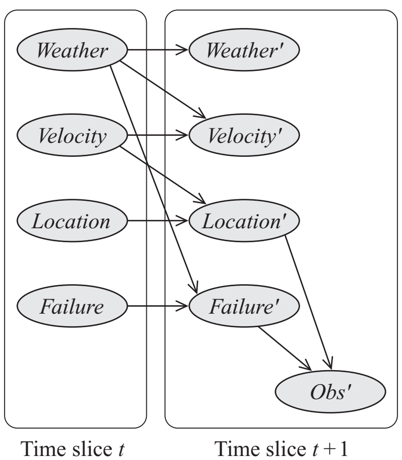
- All variables are *interface* variables except for Obs
  - since we assume that the sensor observation is generated at each point independently given other variables



(a) $\mathcal{B}_{\rightarrow}$

2-Time slice Bayesian Network (2-TBN)

- A 2-Time slice BN for a process over χ is a *conditional* Bayesian Network over χ' given χ_I, where $\chi_I \subseteq \chi$ is a set of interface variables



(a) $\mathcal{B}_{\rightarrow}$

χ={*Weather, Velocity, Location, Failure, Obs*}

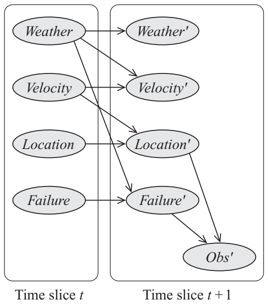
χ'={*Weather', Velocity', Location', Failure', Obs'*}

Interface Variables:
χ_I={*Weather, Velocity, Location, Failure*}

$$\boxed{P(\chi' \mid \chi_I)}$$

Conditional Bayesian Network

- In a conditional Bayesian network only variables in χ' have parents or CPDs



(a) $\mathcal{B}_{\rightarrow}$

$\chi' = \{Weather', Velocity', Location', Failure', Obs'\}$

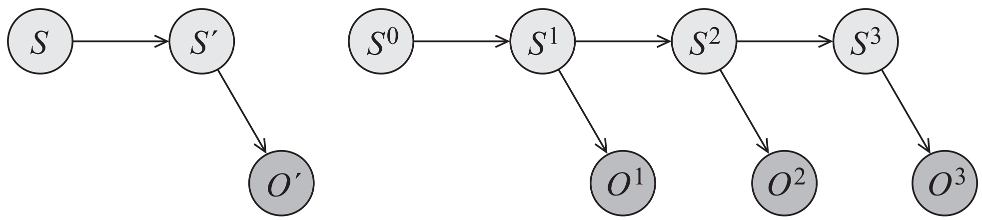
- Interface variables χ_I are those variables whose values at time t have a direct effect on variables at time $t+1$ $\chi_I = \{Weather, Velocity, Location, Failure\}$
  - Thus only variables in χ_I can be parents of variables in χ'
- 2-TBN represents the conditional distribution

$$\boxed{P(\chi' \mid \chi) = P(\chi' \mid \chi_I) = \prod_{i=1}^{n} P(\chi_i' \mid Pa_{\chi_i'})}$$

Example 2-TBN

Simplest nontrivial DBN is a HMM

Single state variable S and single observation variable O



(a) (b)

(a) The 2-TBN for a generic HMM, (b) the unrolled DBN for four time slices

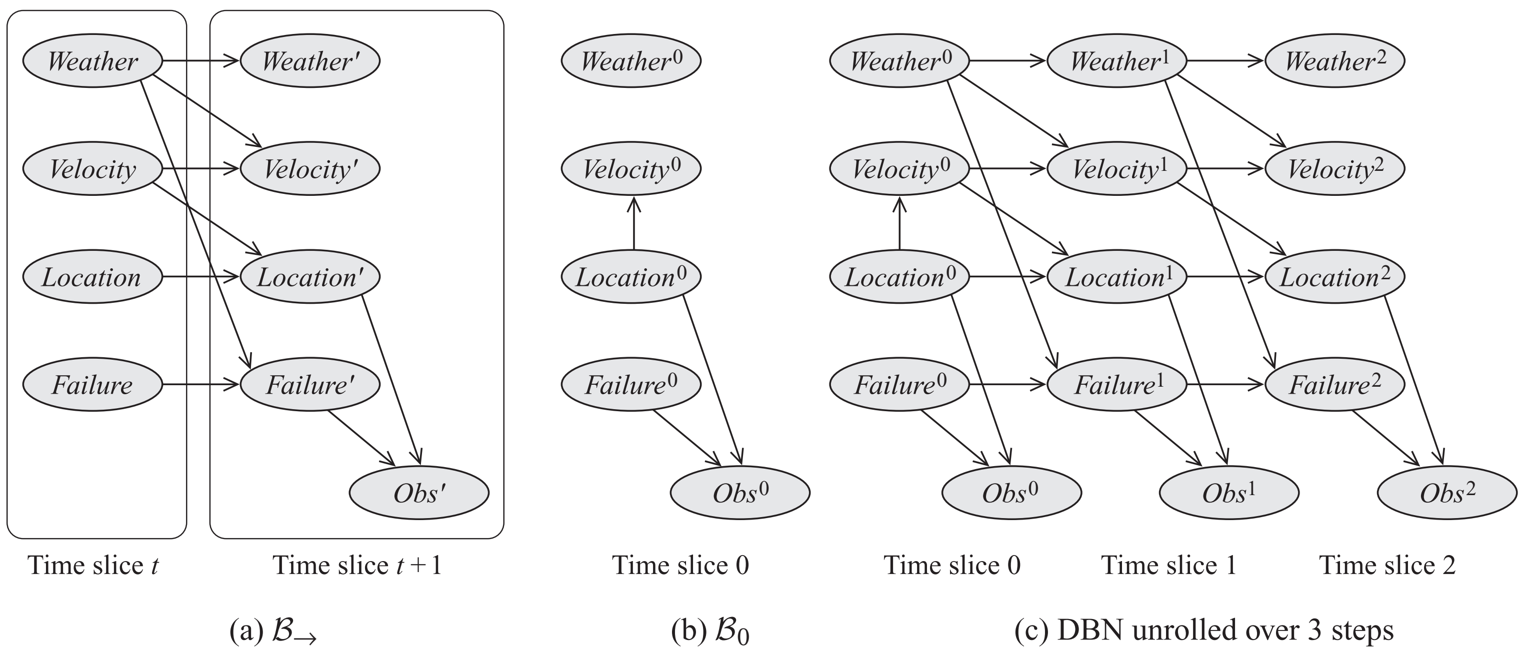
$\chi=\{S, O\}$ $\chi'=\{S', O'\}$ $\chi_I=\{S\}$

$$\boxed{\begin{aligned} &B_{\rightarrow} : P(\chi' \mid \chi_I) = P(S' \mid S) \\ &B_0 : P(S)P(O \mid S) \end{aligned}}$$

Definition of DBN

- A dynamic Bayesian network is a pair $(\mathcal{B}_0, \mathcal{B}_\rightarrow)$
  - $\mathcal{B}_0$ is a Bayesian Network over $\chi^{(0)}$ representing the initial distribution over states
  - $\mathcal{B}_\rightarrow$ Is a 2-TBN for the process
- For any desired time span $T \geq 0$ the distribution over $\chi^{(0:T)}$ is defined as a unrolled Bayesian network where for any $i=1,..,n$:
  - The structure and CPDs of $X_i^{(0)}$ are the same as those for X_i in $\mathcal{B}_0$
  - The structure and CPD of $X_i^{(t)}$ for $t \geq 0$ are the same as those for X_i' in $\mathcal{B}_\rightarrow$

DBN for monitoring a vehicle



(a) The 2-TBN
(b) the time 0 network
(c) resulting unrolled DBN over three time slices

DBN as a compact representation

- A DBN can be viewed as a compact representation from which we can generate an infinite number of Bayesian networks
  - One for every $T > 0$

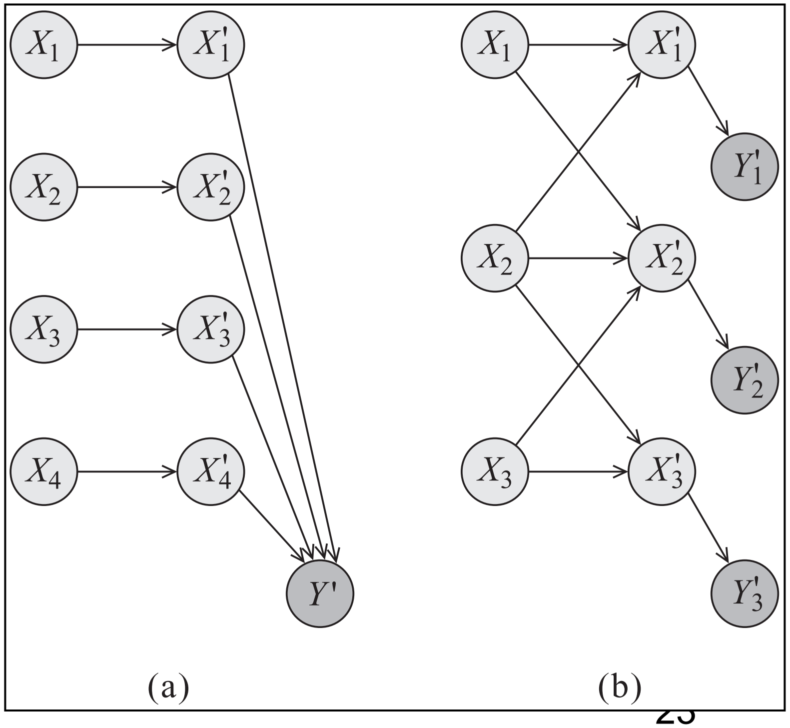
Classes of DBNs from HMMs

(a) A factorial HMM

- 2-TBN has the structure of chains $X_i \rightarrow X_i'$ $(i=1,..n)$
- With a single observed variable Y'
- Ex: several sources of sound through a microphone

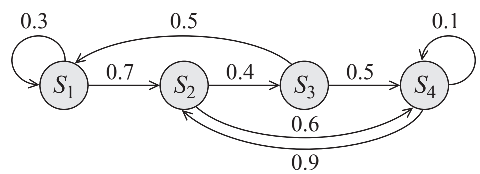
- ## (b) A coupled HMM
  - Also a set of chains X_i
  - Each chain is an HMM with a private observation Y_i
  - Ex: monitoring temperature in a *building* for fire alarms

    X_i is hidden temp of *room*,
    Y_i is sensor reading
    Adjacent room temps interact



(a) (b)

State observation models



- Alternative way of thinking about a temporal process
- State evolves naturally on its own
- Our observation of it is a separate process
- Separates out system dynamics from observation model

State observation model

- Separates out dynamics of system from our ability to sense it
- Two independence assumptions
  1. State variables evolve in a Markovian way
     $(X^{(t+1)} \perp X^{(0:t-1)} \mid X^{(t)})$
  2. Observation at time t are conditionally independent given entire sequence
     $(O^{(t)} \perp X^{(0:t-1)}, X^{(t+1:Inf)} \mid X^{(t)})$
- View model as having two components:
  1. transition model $P(X'|X)$ and
  2. observation model $P(O|X)$

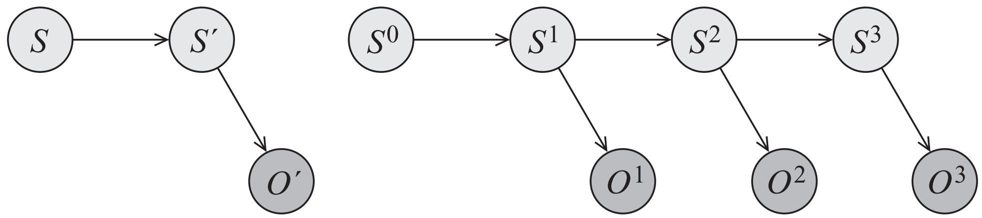
Converting 2-TBN to State-observation

- Any 2-Time slice Bayesian Network can be converted to a state observation representation
  - For any observed variable Y (that does not already satisfy structural restrictions) introduce new variable Y' whose only parent is Y.
  - View Y as being hidden and interpret observations of Y as observations on Y'
    - In effect Y' is a perfectly reliable sensor of Y
- While the transformed network is probabilistically equivalent it obscures independence properties

Applications of State Observation Models

- Hidden Markov Models
- Linear Dynamical Systems

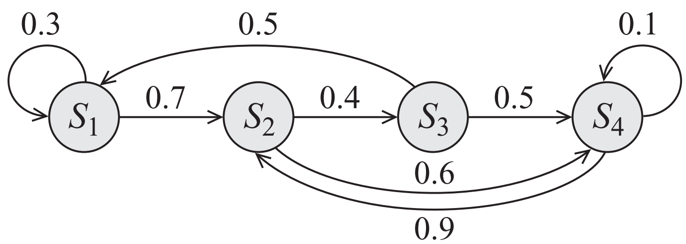
HMMs



(a) (b)

- Transition model $P(S'|S)$ is assumed to be sparse with many possible transitions having zero probability
- Different graphical notation, generally cyclic
- Use representation in which nodes represent different states of the system

HMM transition graphs are very different from PGMs



	s1	s2	s3	s4
s1	0.3	0.7	0	0
s2	0	0	0.4	0.6
s3	0.5	0	0	0.5
s4	0	0.9	0	0.1

- Nodes are states or possible values of the state variables
- Edges represent transitions between states, or entries in the CPDs

HMMs for Speech Recognition

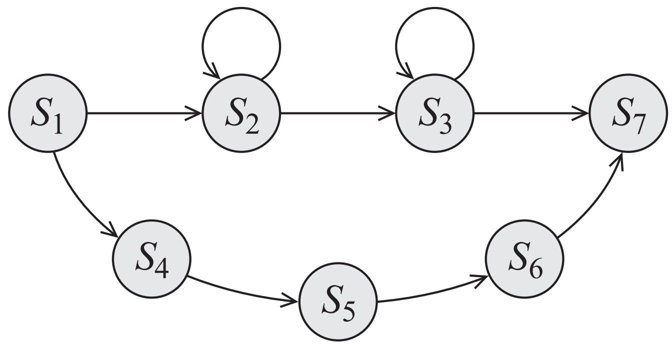
- Three distinct layers

1. Language Model:
   – generates sentences as sequences of words
2. Word Model:
   – described as a sequence of phonemes /p//u//sh/
3. Acoustic model:
   – shows progression of the acoustic signal through a phoneme

Language Model

- Probability distribution over sequences of words Bigram model
  - Markov model defined via distributions $P(W_i \mid W_{i-1})$ for the i^{th} word in sequence
  - Does not take into account position in sentence
    - $P(W_i/W_{i-1})$ is the same for all i
- Trigram model
  - Model distributions as $P(W_i \mid W_{i-1}, W_{i-2})$
- Although naiive this model works well
  - Due to large amounts of training data without manual labeling

Phoneme Model

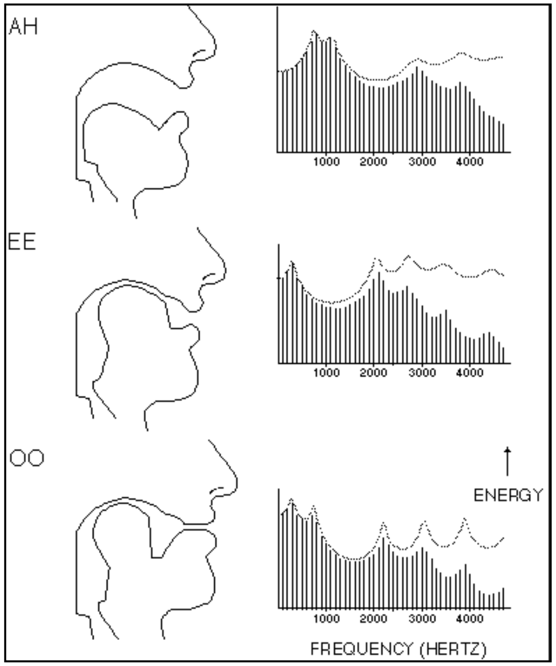


A phoneme-level HMM for a complex phoneme

- Basic phonetic units corresponding to distinct sounds
  - P vs. B
  - Sound is breathy, aspirated, nasalized, and more
  - International phonetic Alphabet has 100 phonemes

Acoustic Level

- Signal segmented into short time frames (around 10-25ms).
- A phoneme lasts over a sequence of these partitions
- Different acoustics for beginning, middle and end of a phoneme
  - Thus a HMM at the phoneme level
  - Observation represents features extracted from acoustic signal
    - Features discretized into bins or a GMM



Combining models with hierarchical HMM

- Three models (language, phoneme and acoustic) combined in a huge hierarchical HMM
  - Defines a joint probability distribution over words, phonemes and basic acoustic units
- In bigram model, states have the form (w,i,j)
  - w =current word, i is a phoneme within that word and j is an acoustic position within that phoneme
- Word HMM has a start state and an end state
  - Each sequence is a trajectory through acoustic HMMs of individual phonemes

Hierarchical HMM to DBN

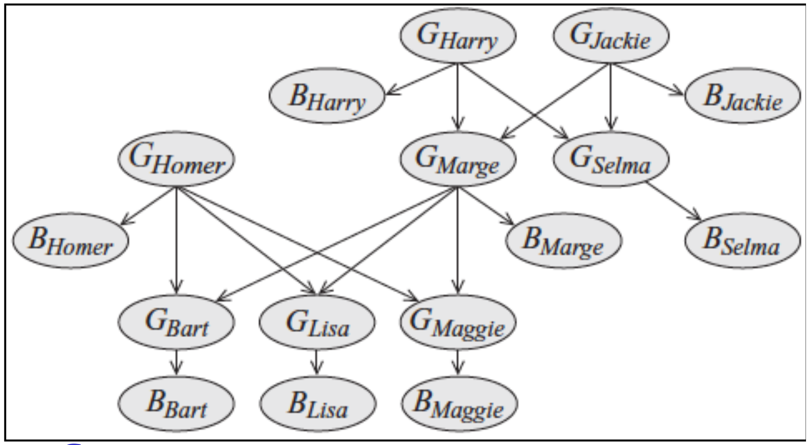
- DBN framework is much more flexible to introduce extensions to the model
- Variables represent different states of different levels of hierarchy (word, phoneme, and intraphone state) along with auxiliary variables (to capture control architecture of hierarchical HMM)

Linear Dynamical Systems

- One or more real-valued variables that evolve linearly over time with some Gaussian noise
- Also called Kalman filters
  - After the algorithm used to perform tracking
- A linear dynamical system can be viewed as a DBN where the variables are all continuous and all the dependencies are linear Gaussian

Another template model: Genetics example

- Family tree (pedigree)
  - individuals all with own properties
- PGM encodes joint distribution over properties of all individuals
- Cannot have a single variable-based model
  - Each family has different family tree
  - Yet mechanism to transmit genes are identical



G: Genotype
B: Blood Type