Cmput 466 / 551

Dynamic Belief Networks

Readings:
A tutorial on hidden Markov models and selected applications…
(Rabiner)

R Greiner
University of Alberta

Dynamic Belief Networks

- Foundations
- Markov Chains (Classification)
- Hidden Markov Models (HMM)
  - Learning HMMs
- Kalman Filter
- General: Dynamic Belief Networks (DBN)
- Applications
- Future Work, Extensions, ...

Why Temporal?

So far: Model world at SINGLE time

- Eg, repairing a car
  - (Stochastically) infer state of car from evidence
  - (car-state/evidence does not change during diagnosis)

What about time?

- Eg, treating a diabetic patient
  - Infer state of patient from evidence
    (insulin doses, food intake, blood sugar, ...)
  - Sequence of measurements ...
    - Blood sugar level over time,
    - depending on food + insulin

$\Rightarrow$ to determine state at time t ... to decide about Rx
need to know history of measurements
(CHO_1, bg_1, $insulin_1$, CHO_2, bg_2, $insulin_2$, ..., CHO_t, bg_t, $insulin_t$)

- Model: sequence of Random Variables:
  One for each aspect of world, for each point in time

Markovian Models

- In general, X_{t+1} depends on everything earlier: $X_t, X_{t-1}, X_{t-2}, \ldots$
- Markovian means... Future ⊥ Past | Present

Future is independent of the past,
once you know the present.

$$P(X_{t+1} \mid X_t, X_{t-1}, , X_{t-2}, \ldots) = P(X_{t+1} \mid X_t)$$

- Markov Chain: "state" (everything important) is visible

$$P(x_{t+1} \mid x_t, \langle \text{everything earlier} \rangle) = P(x_{t+1} \mid x_t)$$

- Eg: First-Order Markov Chain
  1. Random Walk along x axis, changing x-position ± 1 at each time
  2. Predicting rain

r_1	$P(+r_2 \mid r_1)$
+	**0.7**
−	**0.2**

r_2	$P(+r_3 \mid r_2)$
+	**0.7**
−	**0.2**

- Stationarity:

$$P(\text{rain-Tues} \mid \text{rain-Mon}) = P(\text{rain-Wed} \mid \text{rain-Tues}) = \ldots = P(r_{t+1} \mid r_t)$$

Using Markov Chain, for Classification

- Two classes of DNA...
  different di-nucleotide distribution

$$a^{+}_{i,j} = P(x_i \mapsto x_j \mid +) = p^{+}(X_{t+1} = x_j \mid X_t = x_i)$$

$$a^{-}_{i,j} = P(x_i \mapsto x_j \mid -) = p^{-}(X_{t+1} = x_j \mid X_t = x_i)$$

+	A	C	G	T
A	0.180	0.274	0.426	0.120
C	0.171	0.368	0.274	0.188
G	0.161	0.339	0.375	0.125
T	0.079	0.355	0.384	0.182

−	A	C	G	T
A	0.300	0.205	0.285	0.210
C	0.322	0.298	0.078	0.302
G	0.248	0.246	0.298	0.208
T	0.177	0.239	0.292	0.292

- Use this to classify a nucleotide sequence

  $\bar{\mathbf{x}} = \langle$ GATTACACCA... $\rangle$

A: $P(\bar{\mathbf{x}} \mid +) =$

$p^{+}(x_1)\, p^{+}(x_2 \mid x_1)\, p^{+}(x_3 \mid x_2) \ldots p^{+}(x_k \mid x_{k-1}) =$

$\prod_{i=1}^{k} p^{+}(x_i \mid x_{i-1}) = \prod_{i=1}^{k} a^{+}_{x_i \mid x_{i-1}}$

using Markov properties

Using Markov Chain, for Classification

$$a^+_{i,j} = P(x_i \mapsto x_j \mid +) = p^+(X_{t+1} = x_j \mid X_t = x_i)$$

$$a^-_{i,j} = P(x_i \mapsto x_j \mid -) = p^-(X_{t+1} = x_j \mid X_t = x_i)$$

+	A	C	G	T
A	0.180	0.274	0.426	0.120
C	0.171	0.368	0.274	0.188
G	0.161	0.339	0.375	0.125
T	0.079	0.355	0.384	0.182

–	A	C	G	T
A	0.300	0.205	0.285	0.210
C	0.322	0.298	0.078	0.302
G	0.248	0.246	0.298	0.208
T	0.177	0.239	0.292	0.292

- Is $\bar{\mathbf{x}} = \langle ACATTGACCAT \rangle$ in + class?

$P(x \mid +) = p^+(x_1 \mid\ ) \, p^+(x_2 \mid x_1) \, p^+(x_3 \mid x_2) \ldots p^+(x_k \mid x_{k-1})$

$= p^+(A) \quad p^+(C \mid A) \quad p^+(A \mid C) \ldots \; p^+(T \mid A)$

$= 0.25 \times 0.274 \times 0.171 \times \ldots \times 0.355$

$P(x \mid -) = p^-(x_1 \mid\ ) \, p^-(x_2 \mid x_1) \, p^-(x_3 \mid x_2) \ldots p^-(x_k \mid x_{k-1})$

$= p^-(A) \quad p^-(C \mid A) \quad p^-(A \mid C) \ldots p^-(T \mid A)$

$= 0.25 \times 0.205 \times 0.322 \times \ldots \times 0.239$

- Pick larger: "+" if $p(x \mid +) > p(x \mid -)$

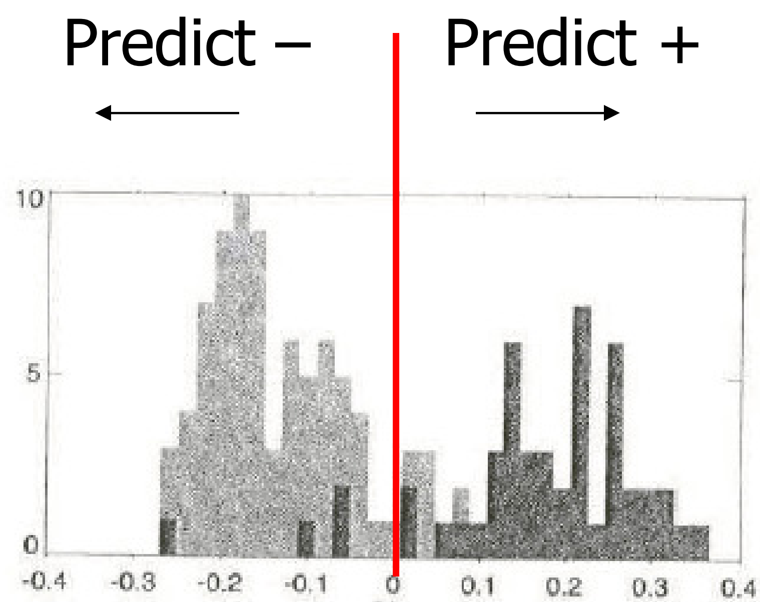
Results (Markov Chain)

- $S(x) = \log \frac{P(x|+)}{P(x|-)} = \sum_{i=1}^{k} \log \frac{a^{+}_{x_{i-1},x_i}}{a^{-}_{x_{i-1},x_i}} = \sum_{i=1}^{k} \beta_{x_{i-1},x_i}$

  where $\beta_{i,j} = \log \frac{a^{+}_{i,j}}{a^{-}_{i,j}}$

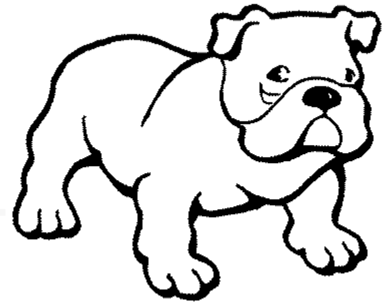
β	A	C	G	T
A	-0.740	0.419	0.580	-0.803
C	-0.913	0.302	1.812	-0.685
G	-0.624	0.461	0.331	-0.730
T	-1.169	0.573	0.393	-0.679

- Results over 48 sequences:



- Here: everything is visible
- Sometimes, can't see the "states"

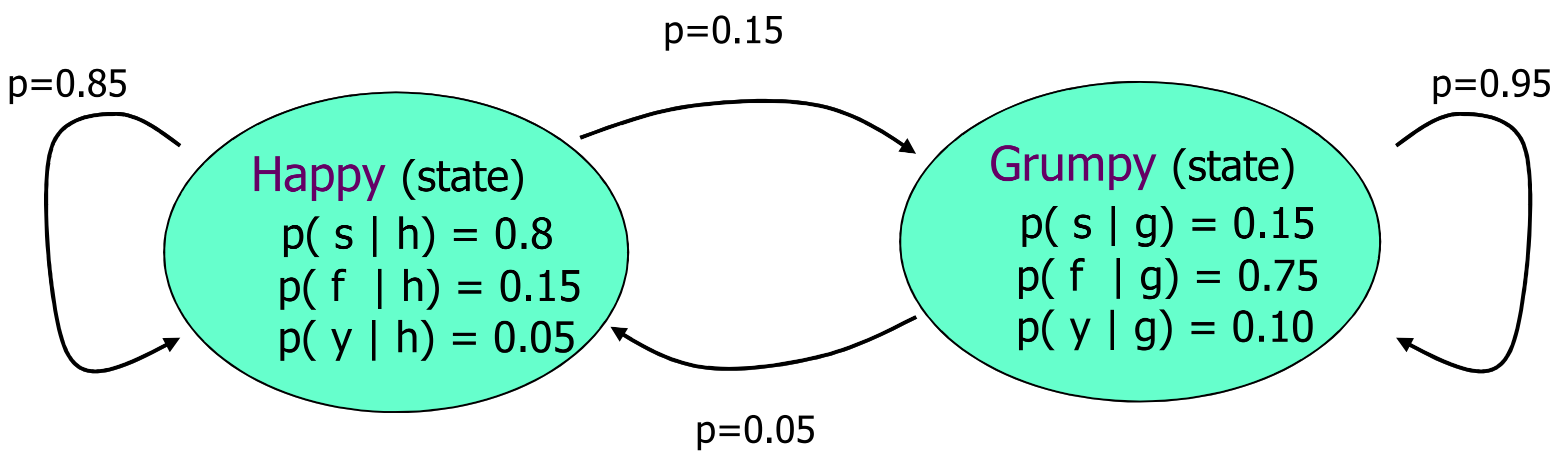
Phydeaux, the Dog



- Sometimes: *Grumpy*
  Sometimes: *Happy*
- But hides emotional state…
  Only observations:
  { *slobbers, frowns, yelps* }

Known Correlations

- State { G,H } to
  Observations {s, f, y}
- State { G,H } on day t to
  state { G,H } on day t+1

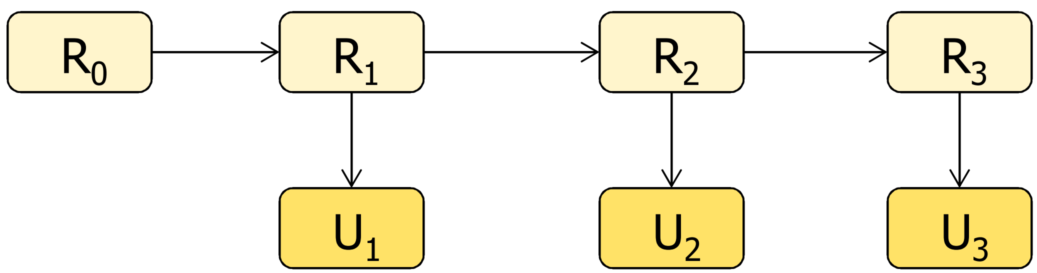


- Challenge: Given observation sequence: ⟨ s, s, f, y, y, f, … ⟩
  what were Phydeaux's states? ??
  ?? ⟨ H, H, H, G, G, G, … ⟩
  ?? ⟨ H, H, G, G, G, H, … ⟩

Umbrella+Rain Situation

- State: $R_t \in \{ +rain, \; -rain \}$
- Observation: $U_t \in \{+umbrella, -umbrella \}$
- Simple (temporal) Belief Net:

r_t	$P(+r_{t+1} \mid r_t)$
+	0.7
–	0.2

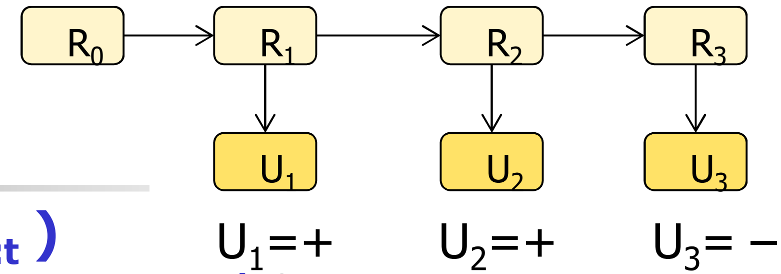


r_t	$P(+u_t \mid r_t)$
+	0.9
–	0.2

- Note: $Umbrella_t$ depends only on $Rain_t$

  $Rain_{t+1}$ depends only on $Rain_t$

HMM Tasks

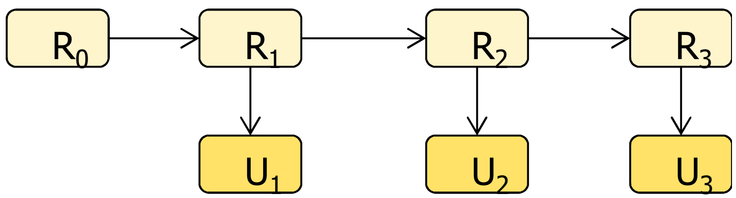


1. Filtering / Monitoring: $\mathbf{P(X_t \mid e_{1:t})}$
   - What is $P(R_3 = + \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use dist'n over current state to make rational decisions
2. Prediction: $\mathbf{P(X_{t+k} \mid e_{1:t})}$
   - What is $P(R_5 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use to evaluate possible courses of actions
3. Smoothing / Hindsight: $\mathbf{P(X_{t-k} \mid e_{1:t})}$
   - What was $P(R_1 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
4. Likelihood: $\mathbf{P(e_{1:t})}$
   - What is $P(U_1 = +, U_2 = +, U_3 = -)$?
   - For comparing different models … classification
5. Most likely expl'n: $\mathbf{argmax_{x1:t}\{ P(x_{1:t} \mid e_{1:t}) \}}$
   - Given $\langle U_1 = +, U_2 = +, U_3 = - \rangle$,
     what is most likely value for $\langle R_1, R_2, R_3 \rangle$?
   - Compute assignments, for DNA, sounds, ...
6. Learning the parameters

Notes

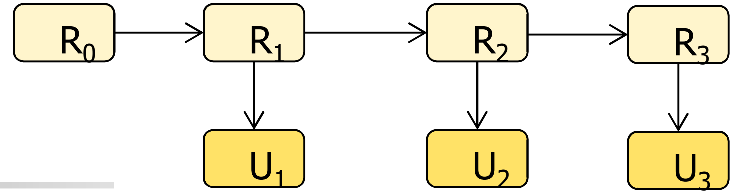
- $e_{1:t} = [e_1, \ldots, e_t]$
  $u_{1:3} = [u_1, u_2, u_3]$
  …

- Can compute P(x) from TREE-Structured belief net, in linear time
  - HMM model is tree-structured



- $P(a \mid b) = \dfrac{P(a,b)}{P(b)} = \dfrac{P(a,b)}{\sum_{a'} P(A=a', b))}$

1. Filtering



$U_1 = +$ $\quad U_2 = +$ $\quad U_3 = -$

- At time 2: have
  - $P(R_2 \mid u_{1:2}) = \langle P(+r_2 \mid +u_1, +u_2), P(-r_2 \mid +u_1, +u_2) \rangle$
  - … then observe $u_3 = -$ … what is $P(R_3 \mid +u_1, +u_2, -u_3)$?

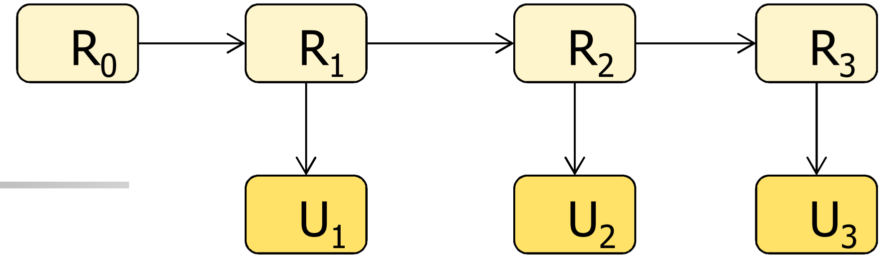
- $$\boxed{P(R_3 \mid u_{1:3})} = \alpha' P(R_3, u_{1:3}) = \alpha' P(R_3, u_{1:2}, u_3)$$
  $$= \alpha P(u_3 \mid R_3, u_{1:2})\ P(R_3 \mid u_{1:2})$$
  $$= \alpha P(u_3 \mid R_3) \quad \boxed{P(R_3 \mid u_{1:2})}$$

$$\alpha' = \frac{1}{P(u_{1:3})}$$

$$\alpha = \frac{P(u_{1:2})}{P(u_{1:3})}$$

- $$\boxed{P(R_3 \mid u_{1:2})} = \sum_{r_2} P(R_3, r_2 \mid u_{1:2})$$
  $$= \sum_{r_2} P(R_3 \mid r_2, u_{1:2})\ P(r_2 \mid u_{1:2})$$
  $$= \sum_{r_2} P(R_3 \mid r_2) \quad \boxed{P(r_2 \mid u_{1:2})}$$

1. Filtering



$U_1 = +$ $\quad U_2 = +$ $\quad U_3 = -$

- At time t:
  - have $P(\, R_t \mid u_{1:t}\,)$
  - … then update based on u_{t+1}
- Distribution of R_{t+1} wrt time $t+1$

$$P(\, R_{t+1} \mid u_{1:t+1}\,) =$$

$$\alpha \underbrace{\boxed{P(\, u_{t+1} \mid R_{t+1}\,)}}_{\text{Emission Prob's}} \sum_{x_t} \underbrace{\boxed{P(R_{t+1} \mid r_t\,)}}_{\text{Transition Prob's}} \underbrace{\boxed{P(\, r_t \mid u_{1:t}\,)}}_{\text{Distribution wrt time } t}$$

- Called "**Forward Algorithm**"

$P(x_t, e_{1:t})$ vs $P(x_t \mid e_{1:t})$

To compute $P(X_t = a \mid e_{1:t})$:

Hidden State: X (was R for Rain)
Observable: e (was u for umbrella)

1. Compute
   $\langle P(X_t=1, e_{1:t}), \ldots, P(X_t=k, e_{1:t}) \rangle$
2. Using this,
   compute $P(e_{1:t}) = \sum_{i=1..k} P(X_t=i, e_{1:t})$
3. For each a, compute
   $$P(X_t=a \mid e_{1:t}) = P(X_t=a, e_{1:t}) \times \frac{1}{P(e_{1:t})}$$

Normalizing constant: $\alpha = \frac{1}{P(e_{1:t})}$

α could be any other term that does not involve X_t ...

Filtering – Forward Algorithm

- Let $f_{1:t} = P(\ X_t \mid e_{1:t}\ )$
  $= \langle\ P(\ X_t = 1 \mid e_{1:t}\ ), \ldots, P(\ X_t = r \mid e_{1:t}\ )\ \rangle$
  $\mathbf{f_{1:t+1}(X_{t+1})} = P(\ x_{t+1} \mid e_{1:t+1}\ )$
  $= \alpha\ P(\ e_{t+1} \mid x_{t+1}\ ) \sum_{x_t} P(X_{t+1} \mid x_t\ )\ \mathbf{f_{1:t}(x_t)}$

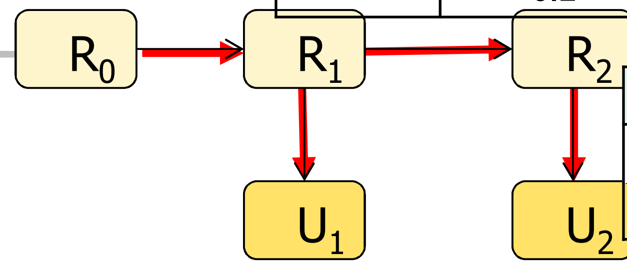
- $f_{1:t+1} = \alpha\ \mathit{Forward}(\ f_{1:t+1}, e_{t+1}\ )$

Detached!

- Update $1{:}t \rightarrow 1{:}t{+}1$ (for discrete state variables):
  Constant time & Constant space!

Forward( )

r_t	$P(+r_{t+1} \mid r_t)$
+	0.7
−	0.2

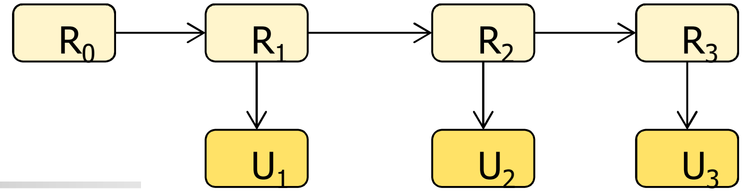


r_t	$P(+u_t \mid r_t)$
+	0.9
−	0.2

$U_1=+$ $\quad$ $U_2=+$

- Given: $P(R_0) = \langle 0.5, 0.5 \rangle$
  Evidence $\langle U_1 = +, U_2 = + \rangle$:
- **Predict state distribution** (before evidence)
  $P(R_1) = \sum_{r_0} P(R_1 \mid r_0)\, P(r_0)$
  $= \langle 0.7, 0.3 \rangle \times 0.5 + \langle 0.2, 0.8 \rangle \times 0.5 = \langle 0.45, 0.55 \rangle$
- **Incorporate** "Day 1 evidence" $+u_1$:
  $P(R_1 \mid +u_1) = \alpha\, P(+u_1 \mid R_1)\, P(R_1)$
  $= \alpha \langle 0.9, 0.2 \rangle .* \langle 0.45, 0.55 \rangle = \alpha \langle 0.405, 0.11 \rangle \approx \langle 0.786, 0.214 \rangle$
- **Predict** (from t = 1 to t = 2, before new evidence)
  $P(R_2 \mid +u_1) = \sum_{r_1} P(R_2 \mid r_1)\, P(r_1 \mid +u_1)$
  $= \langle 0.7, 0.3 \rangle\, 0.786 + \langle 0.2, 0.8 \rangle\, 0.214 \approx \langle 0.593, 0.407 \rangle$
- **Incorporate** "Day 2 evidence" $+u_2$:
  $P(R_2 \mid +u_1, +u_2) = \alpha\, P(+u_2 \mid R_2)\, P(R_2 \mid +u_1) =$
  $\alpha \langle 0.9, 0.2 \rangle .* \langle 0.593, 0.407 \rangle = \alpha \langle 0.534, 0.081 \rangle \approx \langle 0.868, 0.132 \rangle$

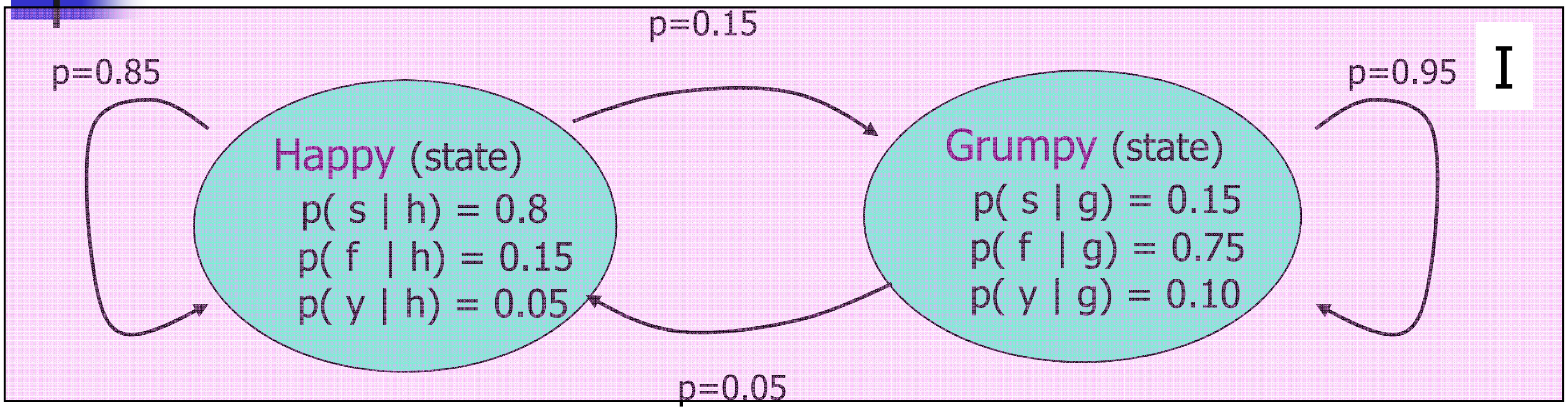
HMM Tasks

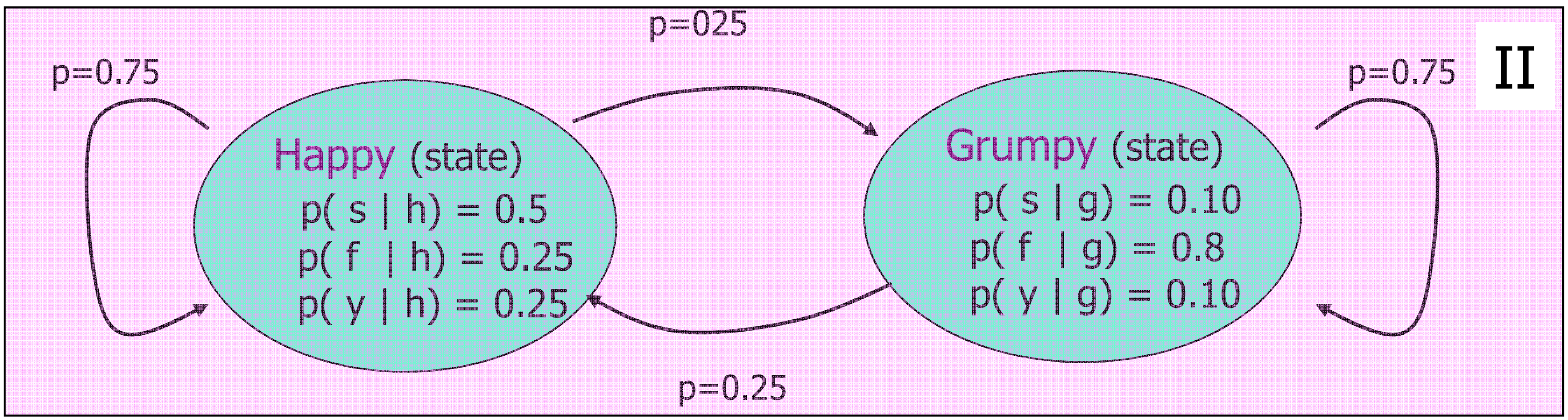


U_1=+ U_2=+ U_3= −

1. Filtering / Monitoring: **$P(X_t \mid e_{1:t})$**
   - What is $P(R_3 = + \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use dist'n. over current state to make rational decisions
2. Prediction: **$P(X_{t+k} \mid e_{1:t})$**
   - What is $P(R_5 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use to evaluate possible courses of actions
3. Smoothing / Hindsight: **$P(X_{t-k} \mid e_{1:t})$**
   - What was $P(R_1 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
4. Likelihood: **$P(e_{1:t})$**
   - What is $P(U_1 = +, U_2 = +, U_3 = -)$?
   - For comparing different models … classification
5. Most likely expl'n: **$\text{argmax}_{x_{1:t}}\{ P(x_{1:t} \mid e_{1:t}) \}$**
   - Given $\langle U_1 = +, U_2 = +, U_3 = - \rangle$,
     what is most likely value for $\langle R_1, R_2, R_3 \rangle$?
   - Compute assignments, for DNA, sounds, ...
6. Learning the parameters

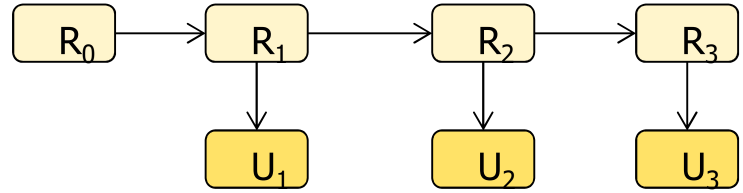
Best Model of Phydeaux?





Challenge: Given observation sequence: **e** = ⟨ s, s, f, y, y, … ⟩

which model of Phydeaux is "correct"?? Compare $P_I($ **e** $)$ vs $P_{II}($ **e** $)$



$U_1 = +$ $\quad$ $U_2 = +$ $\quad$ $U_3 = -$

4. Likelihood

- How to compute **likelihood** $P(\ e_{1:t}\ )$?
- Let $\mathbf{L_{1:t}} = P(\ \mathbf{X_t},\ e_{1:t}\ )$

$$\mathbf{L_{1:t+1}} = P(\ X_{t+1},\ e_{1:t+1}\ ) = \sum_{x_t} P(x_t, X_{t+1}, e_{1:t}, e_{t+1})$$

$$= \sum_{x_t} P(e_{t+1} \mid x_t, X_{t+1}, e_{1:t})\ P(X_{t+1} \mid x_t, e_{1:t})\ P(\ x_t, e_{1:t})$$

$$= \sum_{x_t} P(e_{t+1} \mid \quad X_{t+1} \quad)\ \ P(X_{t+1} \mid x_t \quad)\ P(\ x_t, e_{1:t})$$

$$= P(\ e_{t+1} \mid X_{t+1}\ ) \sum_{x_t} P(X_{t+1} \mid x_t\ )\ \mathbf{L_{1:t}(x_t)}$$

- Note: ≈same *Forward( )* algorithm!!
- To compute actual likelihood:

$$P(\ e_{1:t}\ ) = \sum_{x_t} P(\ X_t = x_t,\ e_{1:t}\ ) = \sum_{x_t} L_{1:t}(x_t)$$

Use HMMs to Classify Words in Speech Recognition

- Use one HMM for each word
  - hmm_j for j^{th} word
- Convert acoustic signal to sequence of fixed duration frames (eg, 60ms)

  (Assumes you know start/end of each word in speech signal)

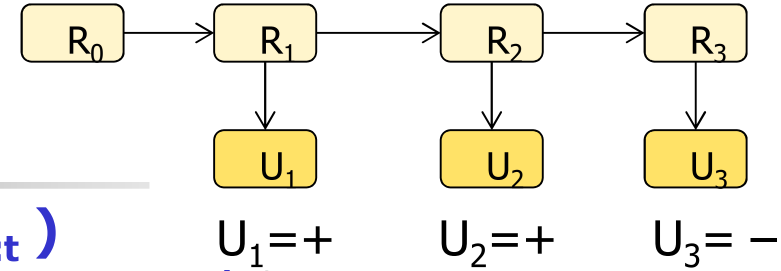
- Map each frame to nearest "codebook" frame (discrete symbol x_t)

1. $e_{1:T} = \langle e_1, \dots, e_n \rangle$

- To classify sequence of frames $e_{1:T}$
  - 1. Compute $P(e_{1:T} \mid hmm_j)$ likelihood $e_{1:T}$ generated by word hmm_j
  - 2. Return $\text{argmax}_j \{ P(e_{1:T} \mid hmm_j) \}$

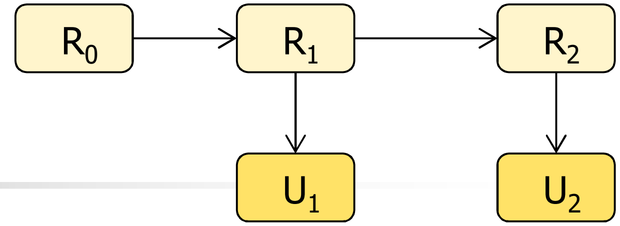
    word#j whose hmm_j gave highest likelihood

HMM Tasks



1. Filtering / Monitoring: $\mathbf{P(\ X_t \mid e_{1:t}\ )}$
   - What is $P(R_3 = + \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use distr. over current state to make rational decisions
2. Prediction: $\mathbf{P(\ X_{t+k} \mid e_{1:t}\ )}$
   - What is $P(R_5 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use to evaluate possible courses of actions
3. Smoothing / Hindsight: $\mathbf{P(\ X_{t\text{-}k} \mid e_{1:t}\ )}$
   - What was $P(R_1 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
4. Likelihood: $\mathbf{P(\ e_{1:t}\ )}$
   - What is $P(U_1 = +, U_2 = +, U_3 = -)$?
   - For comparing different models … classification
5. Most likely expl'n: $\mathbf{argmax_{x1:t}\{\ P(\ x_{1:t} \mid e_{1:t}\ )\ \}}$
   - Given $\langle\ U_1 = +, U_2 = +, U_3 = -\ \rangle$,
     what is most likely value for $\langle\ R_1, R_2, R_3\ \rangle$?
   - Compute assignments, for DNA, sounds, ...
6. Learning the parameters

2. Prediction



U_1=+ U_2=+

- Already have 1 step prediction

  Prediction (from t = 1 to t = 2, before new evidence)

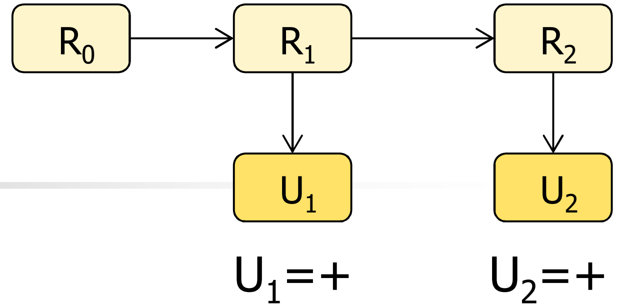
  $P(R_2 \mid +u_1) = \sum_{r1} P(R_2 \mid r_1)\, P(r_1 \mid +u_1) = \ldots \approx \langle 0.627, 0.373 \rangle$

- **Prediction ≡ filtering w/o incorporating new evidence**

  Using transition info, but not observation info

$$P(X_{t+k+1} \mid e_{1:t}) = \sum_{x_{t+k}} P(X_{t+k+1} \mid x_{t+k})\ P(x_{t+k} \mid e_{1:t})$$

2. Prediction



- Already have 1 step prediction

  Prediction (from t = 1 to t = 2, before new evidence)

  $P(R_2 \mid +u_1) = \sum_{r1} P(R_2 \mid r_1)\, P(r_1 \mid +u_1) = \ldots \approx \langle 0.627, 0.373 \rangle$

- **Prediction ≡ filtering w/o incorporating new evidence**

  Using transition info, but not observation info

$$P(X_{t+k+1} \mid e_{1:t}) = \sum_{x_{t+k}} P(X_{t+k+1} \mid x_{t+k})\; P(x_{t+k} \mid e_{1:t})$$

- Converge to stationary distribution **P(Y| e)**

  fixed-point: $P(Y \mid e) = \sum_x P(Y \mid x)\, P(x \mid e)$

  here $\langle 0.5, 0.5 \rangle$

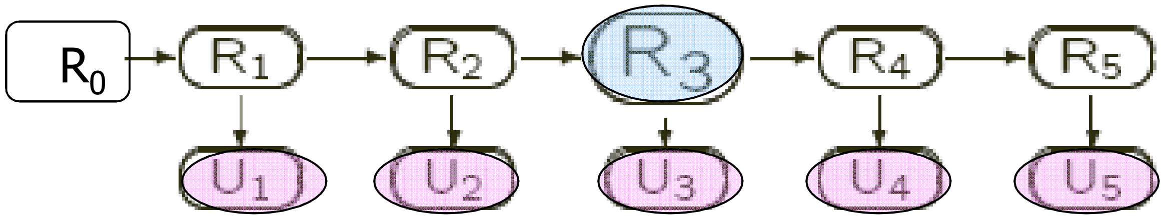
  Mixing time ≈ #steps until reach fixed point

⇒ Prediction meaningless unless k ≪ mixing-time

More “mixing” in transitions

⇒ shorter mixing time, harder to predict future

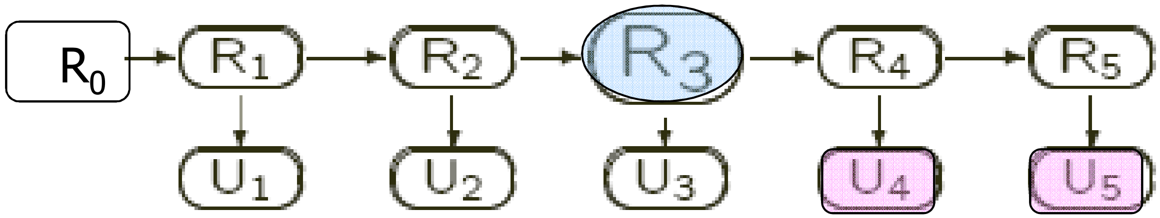
3. Smoothing / Hindsight



- Given $\langle +u_1, +u_2, -u_3, +u_4, -u_5 \rangle$, what is best estimate of r_3

  $P(R_3 \mid +u_1, +u_2, -u_3, +u_4, -u_5)$

3. Smoothing / Hindsight



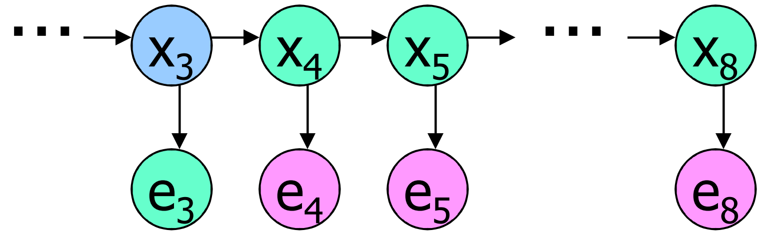
- Given $\langle +u_1, +u_2, -u_3, +u_4, -u_5 \rangle$, what is best estimate of r_3 ?
  $P(R_3 \mid +u_1, +u_2, -u_3, +u_4, -u_5)$
- Let $f_{1:k} = P(X_k \mid e_{1:k})$ $\quad b_{k+1:t} = P(e_{k+1:t} \mid X_k)$

$$
\begin{aligned}
P(X_k \mid e_{1:t}) &= P(X_k \mid e_{1:k}, e_{k+1:t}) \\
&= \alpha\, P(X_k \mid e_{1:k})\ P(e_{k+1:t} \mid X_k, e_{1:k}) \\
&= \alpha\, P(X_k \mid e_{1:k})\ P(e_{k+1:t} \mid X_k) \\
&= \alpha \quad f_{1:k} \quad b_{k+1:t}
\end{aligned}
$$

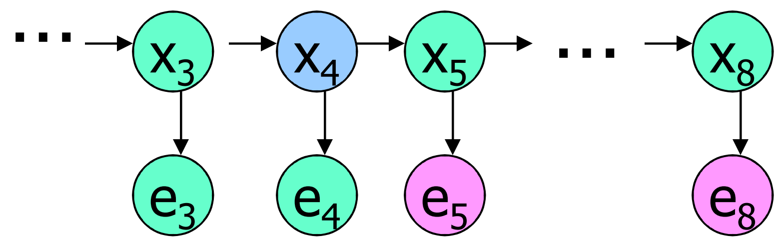
- Recursive computation for $f_{1:k}$ … go forward: *1, 2, 3, …,k*
- Recursive computation for $b_{1:k}$ …go backward: *T, T-1, …,k+1*

Smoothing – Backward Algorithm

$\mathbf{b_{4:8}(x_3)} = P(\ e_{4:8} \mid x_3\ )$



$$
\begin{aligned}
\mathbf{b_{4:8}(x_3)} &= P(\ e_{4:8} \mid x_3\ ) \\
&= \textstyle\sum_{x4} \quad P(\ e_{4:8} \mid x_3, \mathbf{x_4}\ ) \qquad P(\ \mathbf{x_4} \mid x_3\ ) \\
&= \textstyle\sum_{x4} \quad P(\ e_{4:8} \mid \quad x_4\ ) \qquad P(\ x_4 \mid x_3\ ) \\
&= \textstyle\sum_{x4} \quad P(\ \mathbf{e_4, e_{5:8}} \mid x_4\ ) \qquad P(\ x_4 \mid x_3\ ) \\
&= \textstyle\sum_{x4} P(\ e_4 \mid x_4\ )\ \ P(\ e_{5:8} \mid x_4\ )\ \ P(\ x_4 \mid x_3\ ) \\
&= \textstyle\sum_{x4} P(\ e_4 \mid x_4\ ) \quad \mathbf{b_{5:8}(x_4)} \quad P(\ x_4 \mid x_3\ )
\end{aligned}
$$



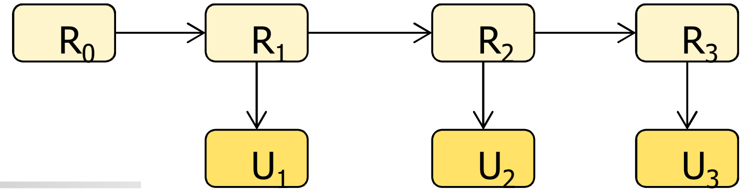
Smoothing – Backward Algorithm

- $\mathbf{b_{k+1:t}(x_k)} = P(\, e_{k+1:t} \mid x_k \,)$
  $= \sum_{x_{k+1}} P(\, ek_{+1:t} \mid xk, xk_{+1}\,)\, P(\, xk_{+1} \mid xk \,)$
  $= \sum_{x_{k+1}} \quad P(\, e_{k+1:t} \mid x_{k+1} \,) \qquad P(\, x_{k+1} \mid x_k \,)$
  $= \sum_{xk+1} P(\, e_{k+1}, e_{k+2:t} \mid x_{k+1} \,)\, P(\, x_{k+1} \mid x_k \,)$
  $= \sum_{xk+1} P(\, e_{k+1} \mid x_{k+1} \,) \;\; P(\, e_{k+2:t} \mid x_{k+1} \,) \;\; P(\, x_{k+1} \mid x_k \,)$
  $= \sum_{xk+1} P(\, e_{k+1} \mid x_{k+1} \,) \quad \mathbf{b_{k+2:t}(x_{k+1})} \quad P(\, x_{k+1} \mid x_k \,)$
- So $b_{k+1:t} = \text{Backward}(\, b_{k+1:t}, e_{k+2:t} \,)$
- Initialize: $b_{t+1:t}(x_t) = P(\, e_{t+1:t} \mid x_t \,) = 1$
- "**Forward-Backward Algorithm**"
  - Just **polytree belief net inference**!
- Fixed-lag smoothing $\langle\, P(\, X_t \mid e_{1:t+k} \,) \,\rangle_t$

Forward-Backward Algorithm

- Inputs:
  - **ev**: vector of evidence values i..t
  - *prior*: $P(X_0)$
- Local vars
  - fv: "forward" msgs for 0..t
  - b: "backward" msgs … initially **1**
  - **sv**: vector of smoothed estimates, 1..t
- **fv**[0] ← *prior*
- for i=1..t do
  - **fv**[*i*] ← Forward(**fv**[*i-1*], **ev**[*i*])
- for *i* = t..1 do
  - **sv**[*i*] ← Normalize(**fv**[*i*] x **b**)
  - **b** ← Backward(**b**, **ev**[*i*])
- return **sv**

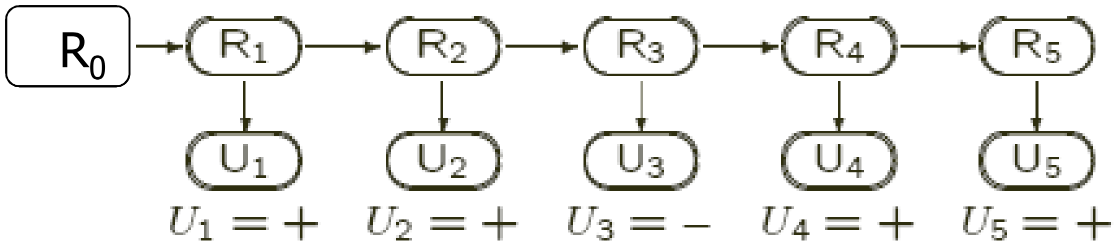
HMM Tasks



U_1=+ U_2=+ U_3= −

1. Filtering / Monitoring: **$P(X_t \mid e_{1:t})$**
   - What is $P(R_3 = + \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use dist'n over current state to make rational decisions
2. Prediction: **$P(X_{t+k} \mid e_{1:t})$**
   - What is $P(R_5 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
   - Use to evaluate possible courses of actions
3. Smoothing / Hindsight: **$P(X_{t-k} \mid e_{1:t})$**
   - What was $P(R_1 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
4. Likelihood: **$P(e_{1:t})$**
   - What is $P(U_1 = +, U_2 = +, U_3 = -)$?
   - For comparing different models … classification
5. Most likely expl'n: **$\text{argmax}_{x_{1:t}}\{ P(x_{1:t} \mid e_{1:t}) \}$**
   - Given $\langle U_1 = +, U_2 = +, U_3 = - \rangle$,
     what is most likely value for $\langle R_1, R_2, R_3 \rangle$?
   - Compute assignments, for DNA, sounds, ...
6. Learning the parameters

5. Most Likely Explanation



- Given $\langle +u_1, +u_2, -u_3, +u_4, +u_5 \rangle$, which is most likely rain-sequence: Perhaps
  - ? $\langle +r_1, +r_2, \mathbf{+r_3}, +r_4, +r_5 \rangle$ but forgot umbrella on day#3?
  - ? $\langle +r_1, +r_2, -r_3, \mathbf{-r_4}, +r_5 \rangle$ but was too cautious on day#4?
  - ? … 2^5 possibilities !
- ? Idea: Just use "**3. Smoothing**" ?

Use argmax_Smoothing for MLE ?

- ? Idea: Use "**3. Smoothing**" ?

  For i = 1..5

     Compute $P(\, R_i \mid \mathbf{u} \,)$

     Let $r_i^* = \text{argmax}_r \{ P(\, R_i = r \mid \mathbf{u} \,) \}$

  Return $\langle r_1^*, \ldots, r_5^* \rangle$

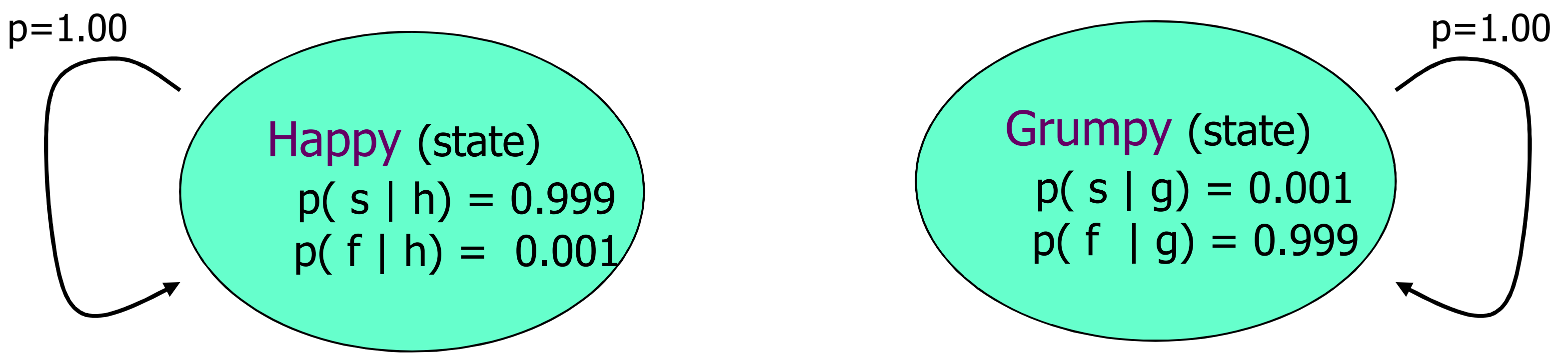
- Most common FIRST name: Mohammad

  Most common LAST name: Wang

  $\Rightarrow$ Most common F+L name: ~~Mohammad Wang~~

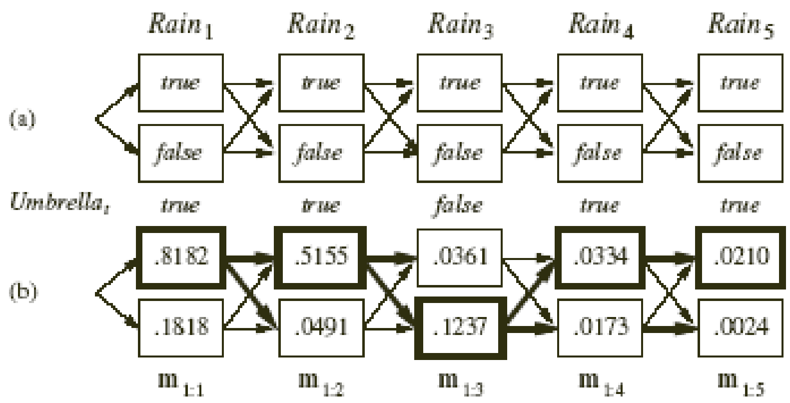
  ?? William Smith ??

Computing MPE …



- Observe ⟨ s, f, s ⟩
- ??Predict ⟨ H, G, H ⟩
- But *0 chance of occurring!!*
- Only possible sequences:
  - ⟨ H, H, H ⟩
  - ⟨ G, G, G ⟩

MLE: Dynamic Program



- Recursively, for each $X_k = x_k$:
  - compute prob of most likely path to each x_k
  - $m_{1:t}(X_t) = \max_{x1,\ldots,xt-1} P(x_1,\ldots,x_{t-1}, X_t \mid e_{1:t})$
- $\mathbf{m_{1:t+1}(X_{t+1})} = \max_{x_1,\ldots,x_t} P(x_{1:t}, X_{t+1} \mid e_{1:t+1})$

  $= P(e_{1:t+1} \mid X_{t+1}) \max_{x_t} [P(X_{t+1} \mid x_t) \max_{x_{1:t-1}} P(x_{1:t-1}, x_t \mid e_{1:t})]$

  $= P(e_{1:t+1} \mid X_{t+1}) \max_{xt} [P(X_{t+1} \mid x_t)\ \mathbf{m_{1:t}(x_t)}]$

MLE – con't

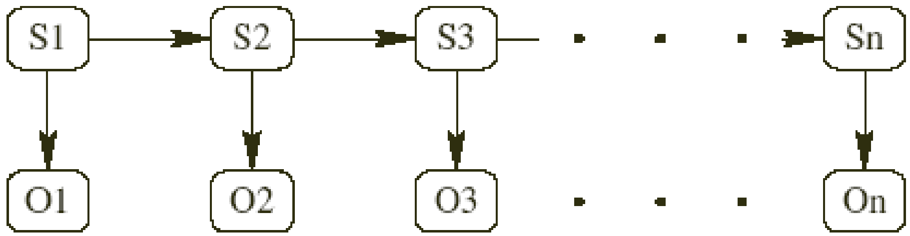
- $\mathbf{m_{1:t+1}} = \max_{x1,\ldots,xt} P(x_{1:t}, X_{t+1} \mid e_{1:t+1})$
  $= P(e_{1:t+1} \mid X_{t+1}) \max_{xt} P(X_{t+1} \mid x_t)\ \mathbf{m_{1:t}}$

- Just like *Filtering* except
  - Replace $f_{1:t} = P(X_t \mid e_{1:t})$
    with $m_{1:t} = \max_{x_{1:t-1}} P(x_{1:t-1}, X_t \mid e_{1:t})$
  - Replace $\sum_{x_t}$ with $\max_{x_t}$

- To recover actual optimal-states x^*_k
  … keep back-pointers!
- **Viterbi Algorithm**

- Linear time, linear space

Most Likely Sequence | DNA



- Observe only output values
  - $\langle$ g c c t a $\rangle$
  - $E_1 = g,\ E_2 = c,\ E_3 = c,\ E_4 = t,\ E_5 = a$
- Want to determine:

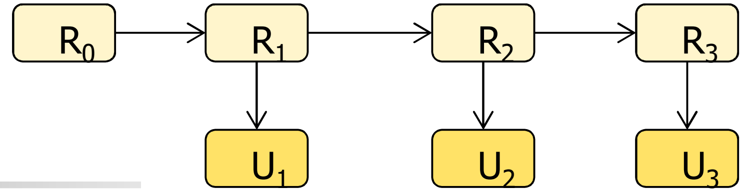
  Most likely sequence of STATES

  - $X_{1:5} = \langle e\ e\ i\ i\ i \rangle$

    $X_1 = e, X_2 = e, X_3 = i, X_4 = i, X_5 = i$

    (e for exon, i for intron)

HMM Tasks



U_1=+ U_2=+ U_3= −

√ 1. Filtering / Monitoring: $\mathbf{P(\ X_t \mid e_{1:t}\ )}$
- What is $P(R_3 = + \mid U_1 = +, U_2 = +, U_3 = -)$?
- Use dist'n over current state to make rational decisions

√ 2. Prediction: $\mathbf{P(\ X_{t+k} \mid e_{1:t}\ )}$
- What is $P(R_5 = - \mid U_1 = +, U_2 = +, U_3 = -)$?
- Use to evaluate possible courses of actions

√ 3. Smoothing / Hindsight: $\mathbf{P(\ X_{t-k} \mid e_{1:t}\ )}$
- What was $P(R_1 = - \mid U_1 = +, U_2 = +, U_3 = -)$?

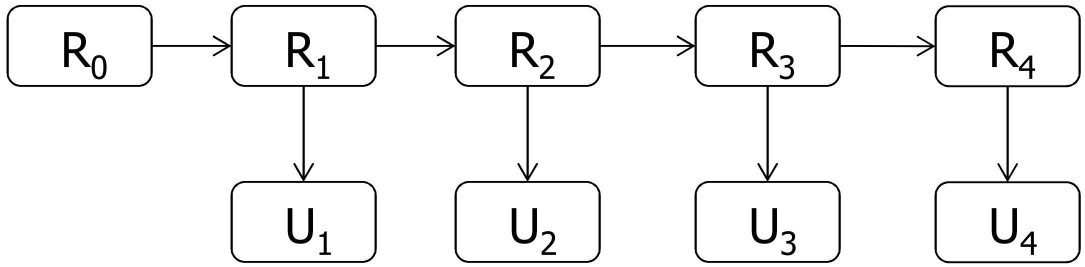
√ 4. Likelihood: $\mathbf{P(\ e_{1:t}\ )}$
- What is $P(U_1 = +, U_2 = +, U_3 = -)$?
- For comparing different models … classification

√ 5. Most likely expl'n: $\mathbf{argmax_{x1:t}\{\ P(\ x_{1:t} \mid e_{1:t}\ )\ \}}$
- Given $\langle U_1 = +, U_2 = +, U_3 = - \rangle$,
  what is most likely value for $\langle R_1, R_2, R_3 \rangle$?
- Compute assignments, for DNA, sounds, ...

6. Learning the parameters

Learning Task

r_t	$P(+r_{t+1} \mid r_t)$
+	$\Theta_{(+r'\mid+r)}$
−	$\Theta_{(+r'\mid-r)}$



r_t	$P(+u_t \mid r_t)$
+	$\boldsymbol{\theta}_{(+u\mid+r)}$
−	$\boldsymbol{\theta}_{(+u\mid-r)}$

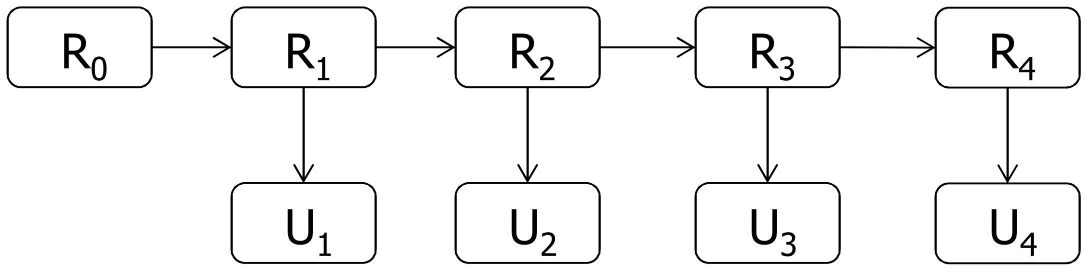
- Given observations:
  [+u1, +u2, −u3, +u4, …] , …
- Find best values for parameters

$$\{\Theta_{(+r'\mid+r)}, \Theta_{(+r'\mid-r)}, \Theta_{(+u\mid+r)}, \Theta_{(+u\mid-r)}\}$$

- How?

Learning Task

r_t	$P(+r_{t+1} \mid r_t)$
+	$\theta_{(+r' \mid +r)}$
–	$\theta_{(+r' \mid -r)}$



r_t	$P(+u_t \mid r_t)$
+	$\theta_{(+u \mid +r)}$
–	$\theta_{(+u \mid -r)}$

- Given *complete* observations (including state):

  $[\, -r_0, (+r_1, +u_1), (-r_2, +u_2), (-r_3, -u_3), (+r_4, +u_4), \ldots], \ldots$

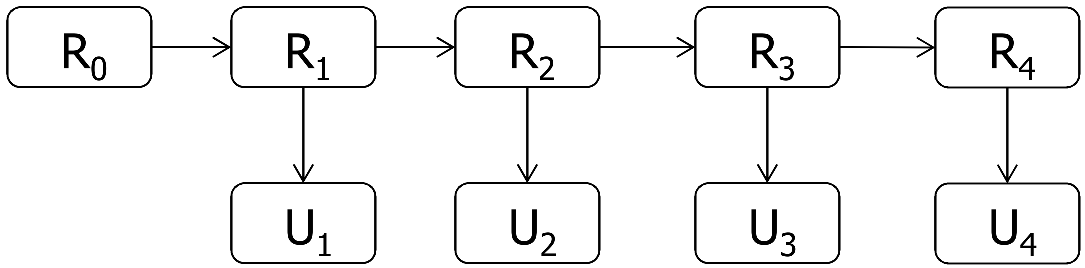
R_t	U_t
+	+
–	+
–	–
+	+

R_t	R_{t+1}
–	+
+	–
–	–
–	+

- Best values for parameters $\{\Theta_{(+r' \mid +r)}, \Theta_{(+r' \mid -r)}, \Theta_{(+u \mid +r)}, \Theta_{(+u \mid -r)}\}$
- Trivial:
  - 2 $+r$ lead to $+u$: so $\theta_{(+u \mid +r)} = 2/2 = 1$
  - 1 $-r$ leads to $+u$, 1 $-r$ leads to $-u$: so $\theta_{(+u \mid -r)} = ½$
  - … similarly for $\Theta_{(+r' \mid +r)}$, $\Theta_{(+r' \mid -r)}$

Learning Task

r_t	$P(+r_{t+1} \mid r_t)$
+	$\Theta_{(+r' \mid +r)}$
−	$\Theta_{(+r' \mid -r)}$



r_t	$P(+u_t \mid r_t)$
+	$\theta_{(+u \mid +r)}$
−	$\theta_{(+u \mid -r)}$

- But given "partial" observations:
  [?, (? ,+u1), (?, +u2), (? , −u3), (? , +u4), …], …

R_t	U_t
?	+
?	+
?	−
?	+

R_t	R_{t+1}
?	?
?	?
?	?
?	?

- If only we knew the values of R_1, R_2, R_3, R_4, …
- Don't know… but can guess…
- Iteratively improve current values of θ:
  - Use MLE values of [R_1, R_2, R_3, R_4]
  - Use DISTRIBUTION over [R_1, R_2, R_3, R_4]

Finding Best Parameters 1: Viterbi

R_t	U_t
?	+
?	+
?	−
?	+

→

R_t	U_t
+	+
−	+
−	−
+	+

1. Guess Initial $\boldsymbol{\theta}^{(0)}$
2. Run **Viterbi Algorithm** to find

   $\mathbf{r}^* = \text{argmax}_{\mathbf{r}}\, P(\, \mathbf{r} \mid \mathbf{u}, \boldsymbol{\theta}^{(0)})$
3. Find ML value of $\boldsymbol{\theta}$ from $\{\, [\, r_i^*, u_i\,]\, \}$
   - $\boldsymbol{\theta}^{(1)} = \text{argmax}_{\boldsymbol{\theta}}\, P(\, \mathbf{r}^* \mid \mathbf{u}, \boldsymbol{\theta}\,)$
   - Factors nicely:

     $\left[\, \boldsymbol{\theta}^{(1)}_{(+\mathbf{u}|+\mathbf{r})}, \boldsymbol{\theta}^{(1)}_{(+\mathbf{u}|-\mathbf{r})} \right]$ depends on $N_{+u,+r}$, $N_{+u,-r}$, …
4. If not done, goto 2.

Finding Best Parameters 2: Baum-Welch

R_t	U_t
?	+
?	+
?	−
?	+

→

R_t	U_t	w_i
+	+	0.3
−	+	0.7
+	+	0.4
−	+	0.6
+	−	0.9
−	−	0.1
+	+	0.1
−	+	0.9

1. Guess Initial $\boldsymbol{\theta}^{(0)}$
2. Run **EM** to find distribution
   $P(\, \mathbf{r} \mid \mathbf{u}, \boldsymbol{\theta}^{(0)})$
3. Find ML value of $\boldsymbol{\theta}$ from $\{\, [\, \mathrm{r_i} \;, \mathrm{u_i}, \;\; w_i \,]\, \}$
   - $\boldsymbol{\Theta}^{(1)} = \operatorname{argmax}_{\boldsymbol{\theta}} P(\, \mathbf{r}^* \mid \mathbf{u}, \boldsymbol{\theta}\,)$
   - Factors nicely:
     $\left[\boldsymbol{\theta}^{(1)}_{(+\boldsymbol{u}|+\mathbf{r})}, \boldsymbol{\theta}^{(1)}_{(+\boldsymbol{u}|-\mathbf{r})}\right]$ depends on $\widehat{N}_{+u,+r}$, $\widehat{N}_{+u,-r}$, …
4. If not done, goto 2.

Comments on Learning Algs

- Viterbi is self-fulfilling …
  - … but fast convergence
- Obvious Bayesian versions … with priors
- Baum-Welch actually PRECEDED EM
  - [Baum&Welch 197**2**]
  - [Dempster, Laird & Rubin 197**7**]
- Can use any techniques for learning Bayesian Network parameters
- Also: if just want $P(u_k \mid u_1, \ldots, u_{k-1})$: can use other methods…

Dynamic Belief Networks

- Foundations
- Markov Chains (Classification)
- Hidden Markov Models (HMM)
- Kalman Filter
- General: Dynamic Belief Networks (DBN)
- Applications
- Future Work, Extensions, ...

Skip

- In 2016: skipped remaining slides …

Kalman Filters

- Tracking a bird in flight, based on (noisy) sensors

  Given observations

  ("estimates" of its position/velocity)

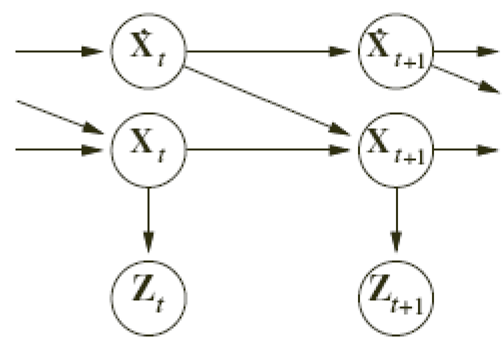
  predict its future position, ...
- X_t = TruePosition @time t

  $\dot{X}_t$ = TrueVelocity @time t

  Z_t = MeasuredPosition @time t
- Observation model: $P(Z_t \mid X_t)$ $\quad Z_t \sim \mathbb{N}(X_t, \sigma_t^2)$

  Transition model: $P(X_{t+1} \mid X_t, \dot{X}_t)$

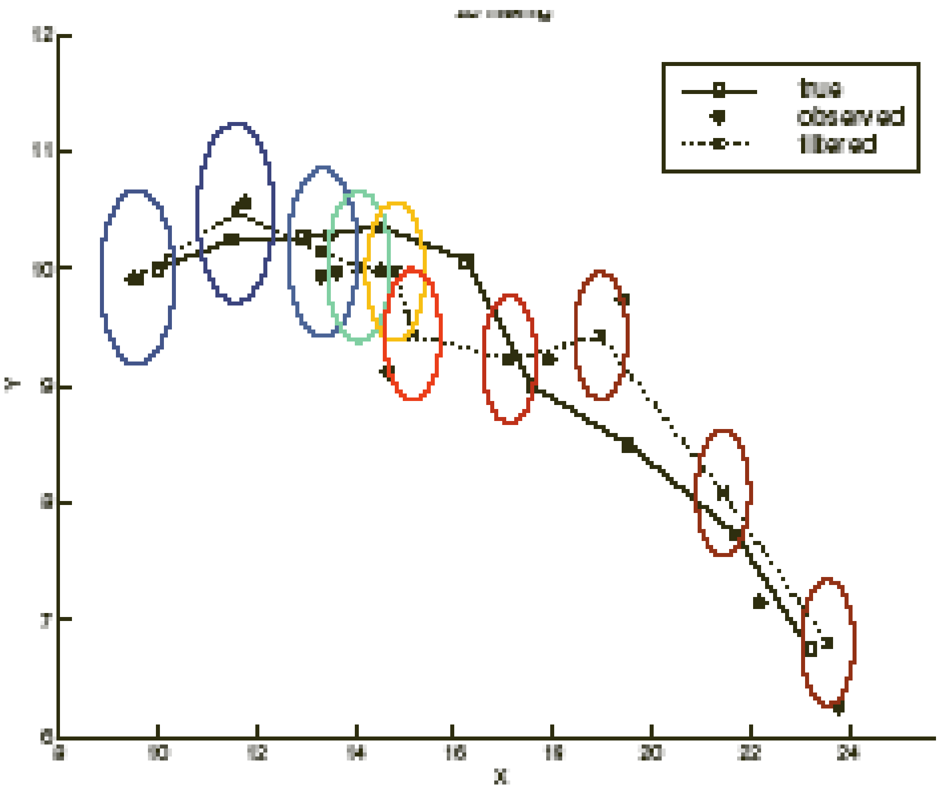
  $X_{t+1} \sim \mathbb{N}(X_t + \dot{X}_t, \sigma_t^2)$

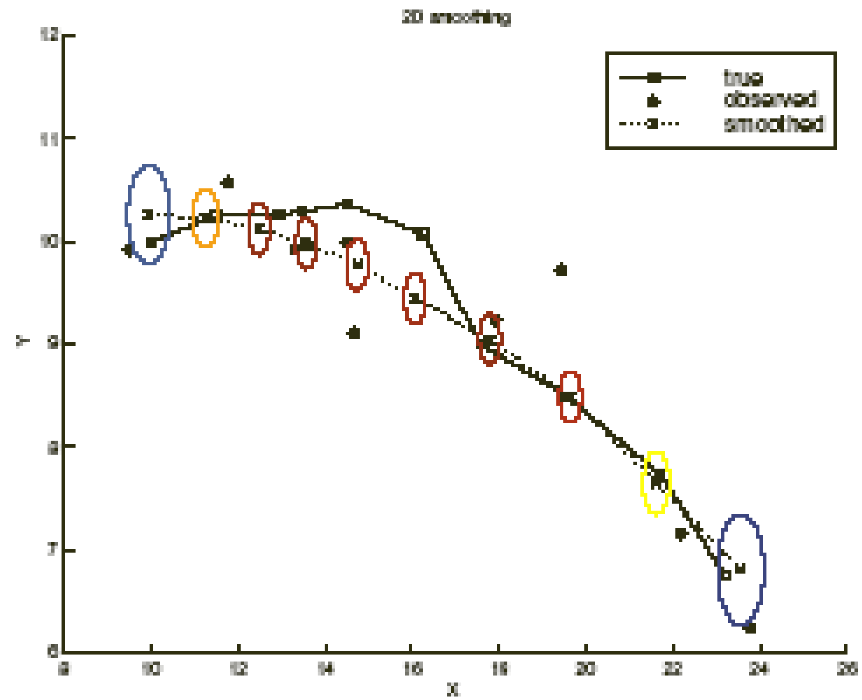


- Everything stays Gaussian!

  … for Filtering, Smoothing, …

Tracking Object in X-Y Plane





Tracking

Smoothing

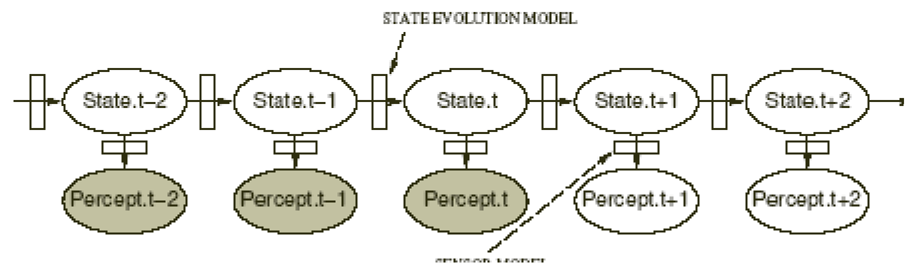
Dynamic Belief Networks

- Foundations
- Markov Chains (Classification)
- Hidden Markov Models (HMM)
- Kalman Filter
- General: Dynamic Belief Networks (DBN)
- Applications
- Future Work, Extensions, ...

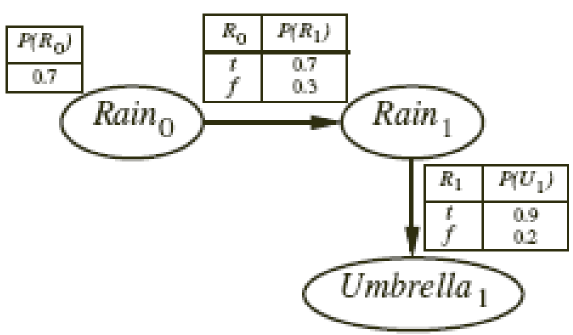
Skip

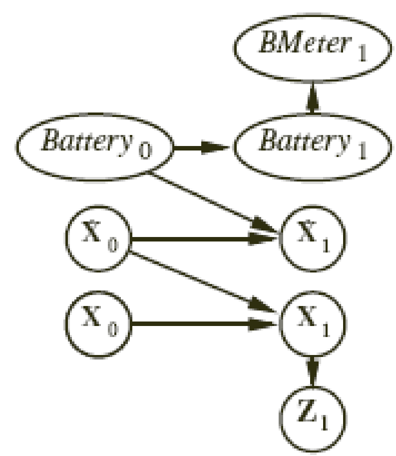
Dynamic Belief Network

- At each time slice:
  - description of state
  - description of observation

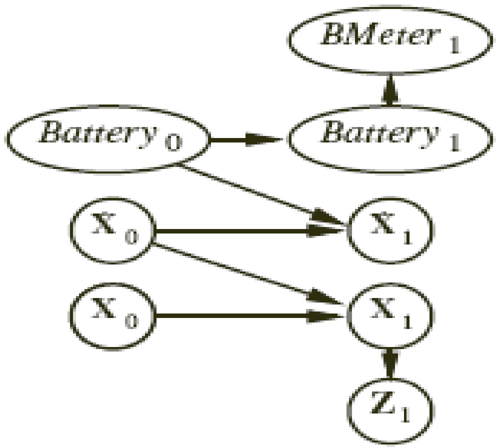


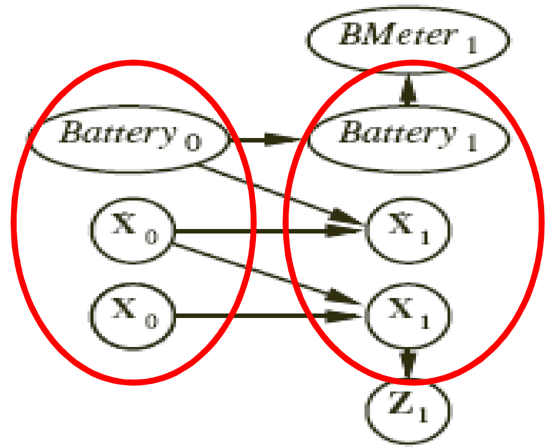
- If 1 var for state, 1 var for obs
  $\Rightarrow$ HMM
- But can have >1 variable for state; >1 for observation!





Advantage of Dynamic BN

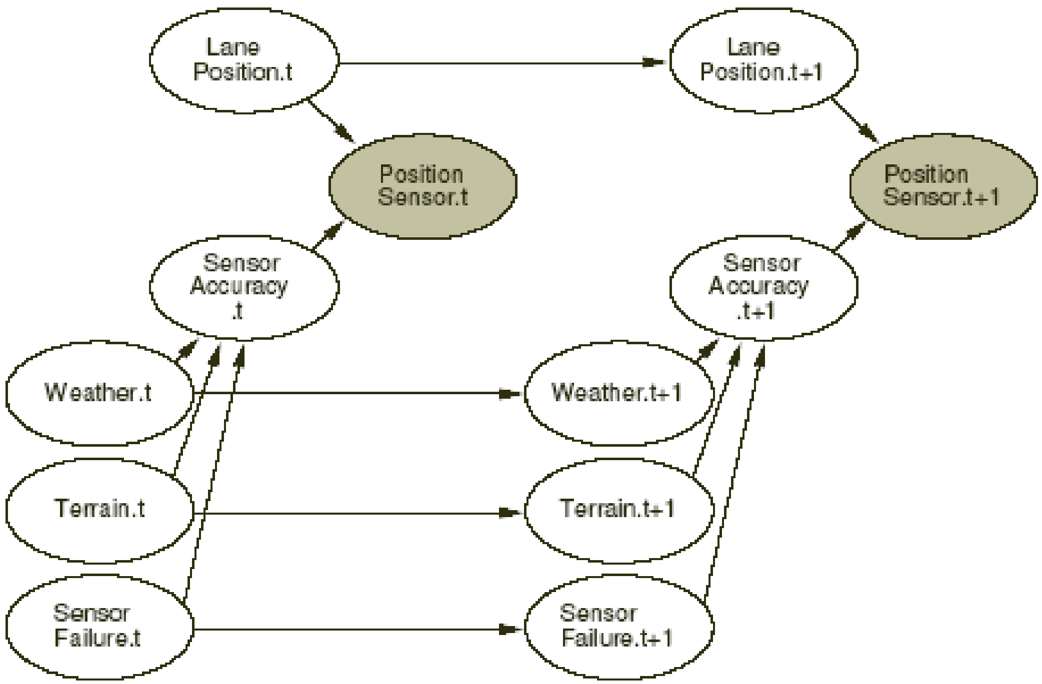




- Why not view DBN as HMM ?
  … just "bundle"
  - the observable variables {BMeter, Z} into 1 meganode
  - the latent variables {X, X', Battery} into 1 meganode
- Answer: Spse |X|=10; |X'|=10; |Battery|=10, |BMeter|=10, |Z|=10
  Now:
  - CPtables: Battery $\rightarrow$ Bmeter: 10x10; X $\rightarrow$ Z: 10x10
    X', $\text{Battery}_t \rightarrow \text{Battery}_{t+1}$: 10x10 x 10; X_t, $X'_t \rightarrow X'_{t+1}$: 10x10 x 10
  - Total: **2,200** values

  As simple HMM:
  - CPtable for Transition Probability: 10x10x10 x 10x10x10 = **1M** !
  - CPtable for Emission Probability: 10x10x10 x 10x10 = **100K**

Representing State as GRAPH of Random Variables

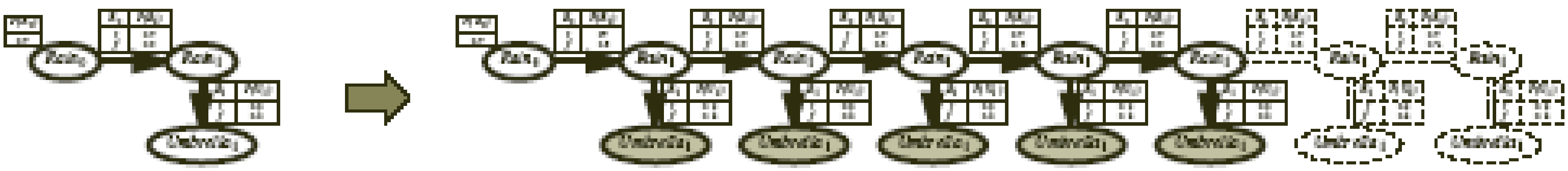


... reduces complexity of representing

$P(X' \mid X, A)$ and $P(E \mid X)$

Inference in DBNs

- As DBN is Belief Net,
  can use std BeliefNet Inference alg
  ... after unrolling



- Filtering

$$\boxed{\begin{aligned} \mathbf{f_{1:t+1}(x_{t+1})} &= P(\, x_{t+1} \mid e_{1:t+1} \,) \\ &= P(\, e_{t+1} \mid x_{t+1} \,) \textstyle\sum_{xt} P(X_{t+1} \mid x_t \,)\, \mathbf{f_{1:t}(x_t)} \end{aligned}}$$

Sums out state variable X_{t-1}

corresponds to Variable Elimination

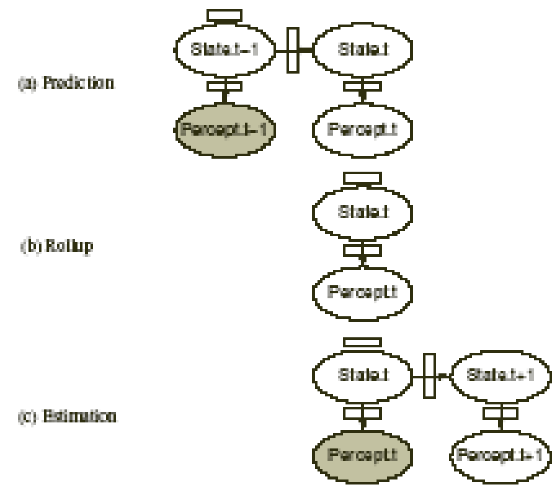
(with this temporal ordering of vars)

Actual DBN Algorithm (Filtering)

- DBN alg: just keep 2 slices in memory

$$\langle X_{t-1}, e_{t-1} \rangle + \langle X_t, e_t \rangle$$

$$f_{1:t+1} = \alpha \text{ Forward}(f_{1:t+1}, e_{t+1})$$



- Constant per-update time, per-update space
- BUT...
  - as Evidence is CHILDREN, parents become COUPLED!

  $\Rightarrow$ constant = $O(d^n)$

  as factor involves all state variables!

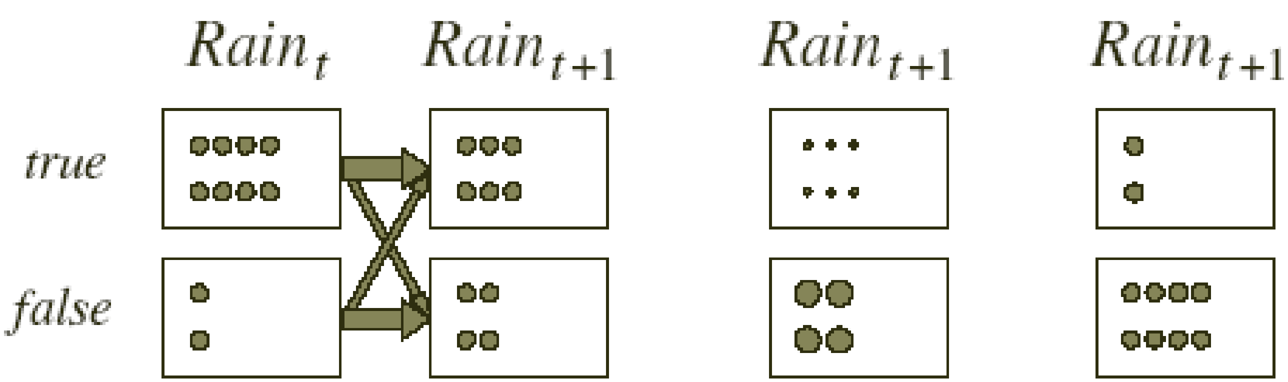
Approximate Algorithms

- Could try...
  - likelihood weighting, MCMC, ...
  - ... but still problems
- Use set of TUPLES themselves as approx'n!
  - Focus on high-probability instances
  - ... tuples ≈ posterior distribution ...
- Particle Filtering

Draw N tuples, $\{d_1^{(0)}, \ldots, d_N^{(0)}\}$, from $P(X_0)$
For $j = 0..\text{Bored}$
  For $i = 1..N$
    Draw x_i from $P(X_{t+1} | d_i^{(t)})$
    Compute weight $w_i = P(e_{t+1} | d_i^{(t+1)})$
  Let $\{ d_i^{(t+1)} \}_i$ be N tuples drawn from $\{ [x_i t+1, w_i] \}_i$

Particle Filtering

Draw N tuples, $\{d_1^{(0)}, \ldots, d_N^{(0)}\}$, from $P(X_0)$
For $j = 0..$Bored
 For $i = 1..N$
  Draw x_i from $P(X_{t+1} | d_i^{(t)})$
  Compute weight $w_i = P(e_{t+1} | d_i^{(t+1)})$
 Let $\{ d_i^{(t+1)} \}_i$ be N tuples drawn from $\{ [x_i t+1, w_i] \}_i$



(a) Propagate (b) Weight (c) Resample

Dynamic Belief Networks

- Foundations
- Markov Chains (Classification)
- Hidden Markov Models (HMM)
- Kalman Filter
- General: Dynamic Belief Networks (DBN)
- Applications
- Future Work, Extensions, ...

Skip

Hierarchical HMMs

- Can construct hierarchy of HMM's:
  - Each Sentence-HMM generates string of word-HMMs
    - each "hidden state" = possible word
  - Each word-HMM generates strings of phoneme-HMMs
    - each "hidden state" = possible phoneme
  - Each phoneme-HMM generates strings of speech frames

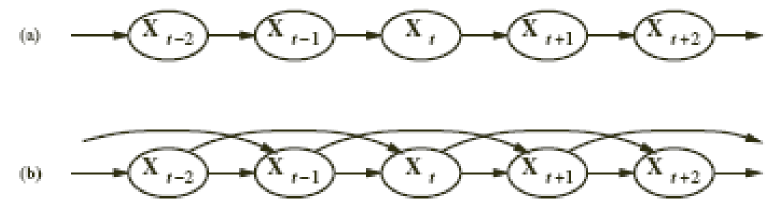
- "Compile" hierarchy into *frame-level HMM* that finds

  whole sentence most likely to have been spoken

- MLE – computed by Viterbi algorithm

Beyond First-Order

- Recall First-Order Markov Chain
  - Random Walk along x axis,
  - changing x-position 1 at each time

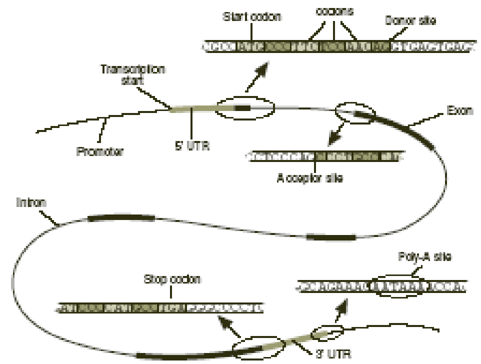


- What if position x_t depends on x_{t-1}, x_{t-2}?
  - (Ie, need velocity, as well as position)
  - 2nd-order Markov Chain

[Can make any process into 1st-order Markov,
by expanding state
Eg, to deal with power being consumed,
could have BatteryLevel in state
. . . in the limit: "state"≡ "all history"]

- Interpolated Markov Model (GLIMMER)

Computational Biology: Find Region of Interest in DNA



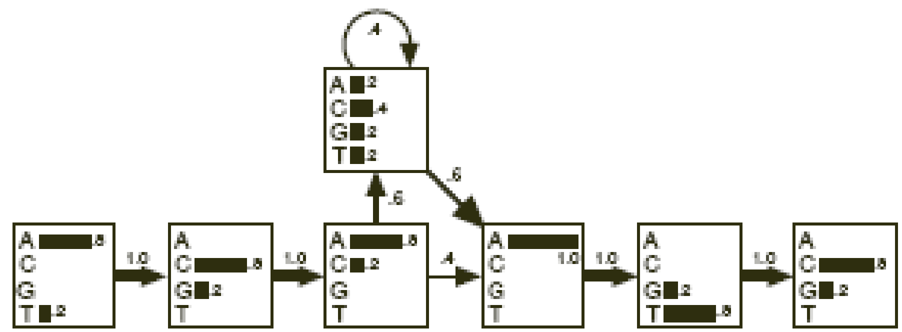
- Segment DNA into
  - Exon vs Intron vs Intergenetic Region
  - StartCodon, DonorSite, AcceptorSite, StopCodon
  - Techniques: NN, DecisionTrees, HMMs
- Identify "motif"
  - "Significant Nucleotide Sequence"
  - Intron/Exon boundary
  - Sites: Promoter, Enhancer,
    Transcription factor binding, Splice cite
  - CRP Binding site (or LexA binding site, or ...)

HMM's in Biological Sequence Data

- Given collection of similar genes
  (eg, same function, but different animals)
  find new genes in other organisms that are similar.
  [Ex: Globins (hemoglobin, myoglobin)]
  Use "**4. Likelihood**" alg
- Given collection of similar genes,
  align them to one another
    (identify where mutations have occurred:
    insertions, deletions, replacements)
  Useful for studying evolution and discovering functionally important parts
  Use "**5. MLE**" alg

Simple Hidden Markov Model

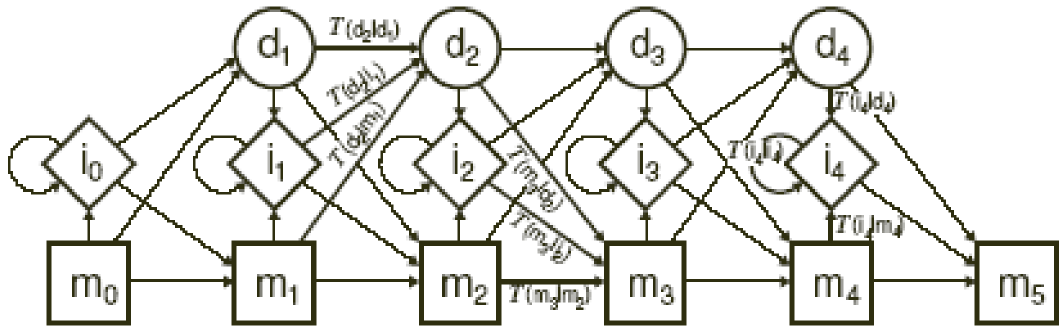
$$\left\{\begin{array}{l}
\langle \mathrm{A}\quad \mathrm{C}\quad \mathrm{A}\quad -\quad -\quad -\quad \mathrm{A}\quad \mathrm{T}\quad \mathrm{G}\rangle \\
\langle \mathrm{T}\quad \mathrm{C}\quad \mathrm{A}\quad \mathrm{A}\quad \mathrm{C}\quad \mathrm{T}\quad \mathrm{A}\quad \mathrm{T}\quad \mathrm{C}\rangle \\
\langle \mathrm{A}\quad \mathrm{C}\quad \mathrm{A}\quad \mathrm{C}\quad -\quad -\quad \mathrm{A}\quad \mathrm{G}\quad \mathrm{C}\rangle \\
\langle \mathrm{A}\quad \mathrm{G}\quad \mathrm{A}\quad -\quad -\quad -\quad \mathrm{A}\quad \mathrm{T}\quad \mathrm{C}\rangle \\
\langle \mathrm{A}\quad \mathrm{C}\quad \mathrm{C}\quad \mathrm{G}\quad -\quad -\quad \mathrm{A}\quad \mathrm{T}\quad \mathrm{C}\rangle
\end{array}\right\}$$



- Each box is “state"
  w/prob of “emitting" a letter
- Transition from state to state
  - Bottom Row: standard “emit a letter"
  - Upper Row: insert “extra” letter

  (After state3, 3/5 of sequences goto “Insert"
  Of 5 transitions from “Insert", 2 goto another insert)
- If no gaps, same as earlier model.

Profile HMM



- Special structure: “profile HMM”
- Main (level 0)

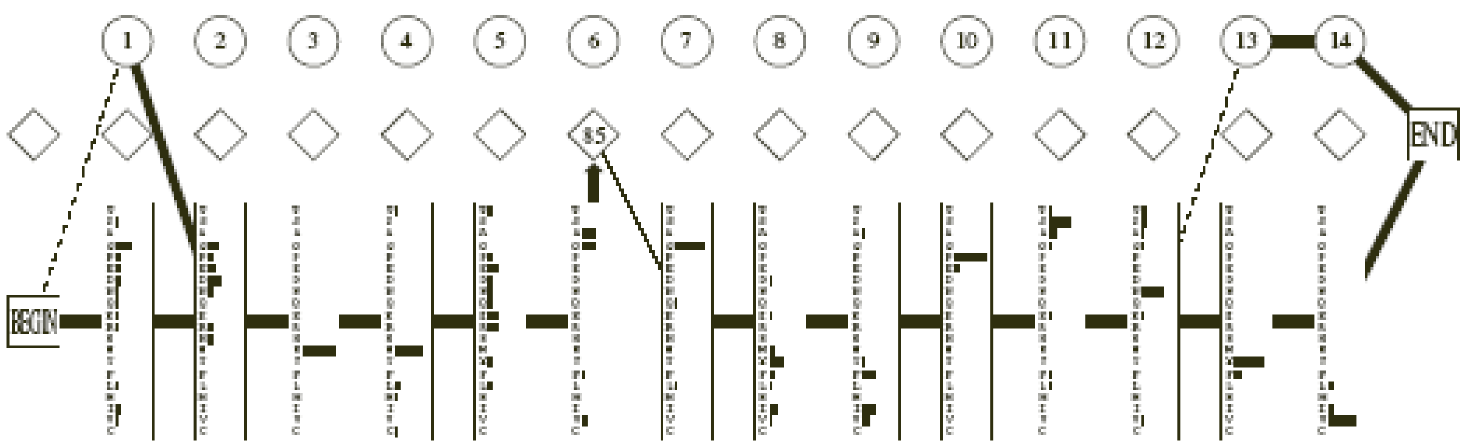
  For “columns” of alignment
- Insert (level 1)

  For highly-variable regions
- Delete (level 2)

  “silent” or “null”

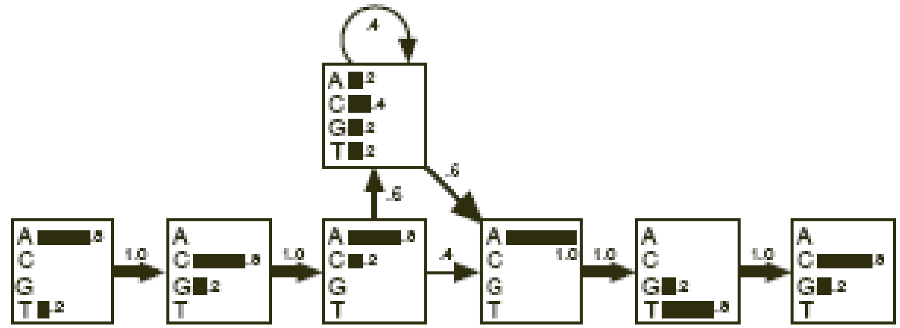
Example

```
GGWWRGdy.ggkkqLWFPSNYV
IGWLNGynettgerGDFPGTYV
PNWWEGql..nnrrGIFPSNYV
DEWWQArr..deqiGIVPSK--
GEWWKAqs..tggeGFIPFNFV
GDWWLArs..sgqtGYIPSNYV
GDWWDAel..kgrrGKVPSNYL
-DWWEArslssghrGYVPSNYV
GDWWYArslitnseGYIPSTYV
GEWWKArslatrkeGYIPSNYV
GDWWLArslvtgreGYVPSNFV
GEWWKAkslsskreGFIPSNYV
GEWCEAqt.kngq.GWVPSNYI
SDWWRVvnlttrqeGLIPLNFV
LPWWRArd.knggeGYIPSNYI
RDWWEFrsktvytpGYYESGYV
EHWWKVkd.algnvGYIPSNYV
IHWWRVqd.rngheGYVPSSYL
KDWWKVev..ndrqGFVPAAYV
VGWMPGlnertrqrGDFPGTYV
PDWWEGel..nggrGVFPASYV
ENWWNGei..gnrkGIFPATYV
EEWLEGec..kgkvGIFPKVFV
GGWWKGdy.gtriqQYFPSNYV
DGWWRGsy..nggvGWFPSNYV
QGWWRGei..ygrvGWFPANYV
GRWWKArr.angetGIIPSNYV
GGWTQGel.ksgqkGWAPTNYL
GDWWEArsn.tgenGYIPSNYV
NDWWTGrt..ngkeGIFPANYV
```



[5] Probability of Sequence wrt HMM

- $P_{HMM}(\langle ACACATC \rangle) =$
  $$\begin{aligned}
  & P(\text{emit } A \mid M_1) \times P(M_1 \to M_2) \times \\
  & P(\text{emit } C \mid M_2) \times P(M_2 \to M_3) \times \\
  & P(\text{emit } A \mid M_3) \times P(M_3 \to I_3) \times \\
  & P(\text{emit } C \mid I_3) \times P(I_3 \to M_4) \times \\
  & P(\text{emit } A \mid M_4) \times P(M_4 \to M_5) \times \\
  & P(\text{emit } T \mid M_5) \times P(M_5 \to M_6) \times \\
  & P(\text{emit } C \mid M_6) \\
  = {} & 0.8 \times 1 \times 0.8 \times 1 \times 0.8 \times 0.6 \times 0.4 \\
  & \times 0.6 \times 1 \times 1 \times 0.8 \times 1 \times 0.8 \\
  \approx {} & 4.7 \times 10^{-2}
  \end{aligned}$$



- Here, unambiguous. . .

  Only consistent path through HMM is

  ⟨ M1, M2, M3, I3, M4, M5, M6⟩

  In general, several possible paths. . .

Recent applications of HMMs

- Proteins
  - detection of bronectin type III domains in yeast
  - a database of protein domain families
  - protein topology recognition from secondary structure
  - modeling of a protein splicing domain
- Gene finding
  - detection of short protein coding regions and analysis of translation initiation sites in Cyanobacterium
  - characterization of prokaryotic and eukaryotic promoters
  - recognition of branch points
- Also
  - prediction of protein secondary structure
  - modeling an oscillatory pattern in nucleosomes
  - modeling site dependence of evolutionary rates
  - for including evolutionary information in protein secondary structure prediction
- Free packages:
  - hmmer – http://genome.wustl.edu/eddy/hmm.html
  - SAM – http://www.cse.ucsc.edu/research/compbio/sam.html

Other Applications

Similar approaches work for analyzing

- Proteins (Amino-Acid sequences)
  - Similar composition, similar function, and …
- Protein Folding
  - Protein sequence of a.a.'s
  - "Tertiary structure" ≡ Complete 3D structure
  - "Secondary structure" ≡ Simpler decomposition
    α-helices, β-sheets, (random) coil
- TEMPORAL sequences
  - weather prediction
  - stock-market forecasting
  - …

Future Research

- Scaling up to handle larger
  { sequences, motifs, DBs }

  Learn...
  - more accurate descriptions
  - in less time (fewer samples, less CPU-time)
  - rep'ns that allow more efficient computation
- Exploiting other information
  - facts about a.a.'s (hierarchy?)
  - structural information
  - ...

Summary

- To model temporal events
  - Use rv X_t to model X at time t
- Markov Property:

  $$P(\mathbf{X}_{t+1} \mid \mathbf{X}_t, \mathbf{X}_{t-1}, \ldots) = P(\mathbf{X}_{t+1} \mid \mathbf{X}_t)$$
- Hidden Markov Model:
  - Emission $P(E_t \mid X_t)$; Transition $P(X_{t+1} \mid X_t)$
  - Efficient (linear time!) to predict …
    - Current state (filtering)
    - Previous state (smoothing)
    - Future state (prediction)
    - Most likely explanation (Viterbi)
- Dynamic Belief Nets – extension of HMM
  … mixing …
- Uses: Speech recognition; Tracking; BioInformatics, …