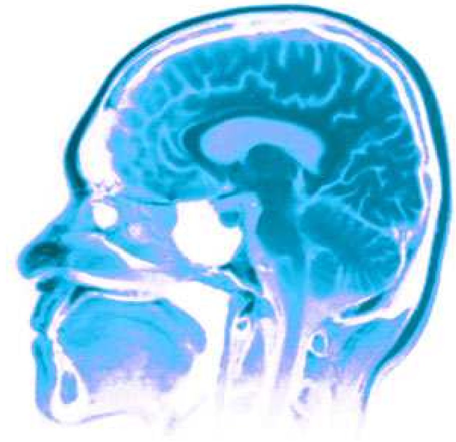
CPSC340



Hidden Markov Models

Nando de Freitas
September, 2012
University of British Columbia

Outline of the lecture

This lecture is devoted to the problem of inference in probabilistic graphical models (aka Bayesian nets). The goal is for you to:

- ❑ Practice marginalization and conditioning.
- ❑ Practice Bayes rule.
- ❑ Learn the HMM representation (model).
- ❑ Learn how to do **prediction** and **filtering** using HMMs

Assistive technology

Assume you have a little robot that is trying to ***estimate*** *the* ***posterior*** *probability that you are* ***happy*** *or* ***sad****, given that the robot has observed whether you are* ***watching Game of Thrones (w)****,* ***sleeping (s)****,* ***crying (c)*** *or* ***face booking (f)****.*

Let the ***unknown state*** *be* ***X=h*** *if you're happy and* ***X=s*** *if you're sad.*

Let ***Y*** *denote the* ***observation****, which can be* ***w****,* ***s****,* ***c*** *or* ***f*** *.*

We want to answer queries, such as:

$P(X=h|Y=f)$?

$P(X=s|Y=c)$?

Assistive technology

Assume that an expert has compiled the following ***prior*** *and* ***likelihood*** *models:*

$P(x|y)$?

$P(x) =$

	X	
	s	H
	0.2	0.8

$P(X=s) = 0.2$

$P(y|x) =$

		y			
		w	s	c	f
x	s	0.1	0.3	0.5	0.1
	H	0.4	0.4	0.2	0

$$\frac{.32}{.32+.02} = 0.94$$

$$P(x=H|y=w) = \frac{\overset{0.4}{P(y=w|x=H)}\,\overset{0.8}{P(x=H)}}{\underset{0.4}{P(y=w|x=H)}\,\underset{0.8}{P(x=H)} + \underset{0.1}{P(y=w|x=s)}\,\underset{0.2}{P(x=s)}}$$

Assistive technology

*But what if instead of an absolute prior, what we have instead is a temporal (**transition prior**). That is, we assume a **dynamical system***

X_1 \ X_2	S	H
S	0.99	0.01
H	0.1	0.9

$$P(X_2 = s \mid X_1 = s) = 0.99$$

init $P(X_0) = \boxed{0.8 \mid 0.2}$

Given a history of observations, say $\underline{Y_1=w}$, $\underline{Y_2=f}$, $\underline{Y_3=c}$, *we want to compute the posterior distribution that you are happy at step 3. That is, we want to estimate:*

$$\boxed{P(X_3=h \mid Y_1=w, Y_2=f, Y_3=c)}$$

Clearly, to know if you're happy when crying it helps to know if the sequence of observations is **wcw** or $\overset{Y_1 Y_2 Y_3}{\textbf{ccc}}$.

Dynamic model

In general, we assume we have an ***initial*** *distribution* $\boldsymbol{P(X_0)}$*, a* ***transition*** *model* $\boldsymbol{P(X_t \mid X_{t-1})}$*, and an* ***observation*** *model* $\boldsymbol{P(Y_t \mid X_t)}$.

✓ $P(x_0) =$

S	H
.2	.8

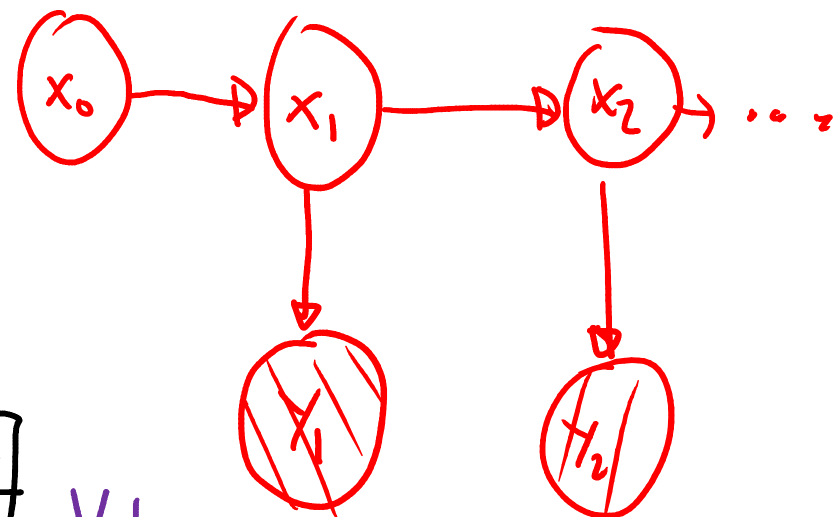
$P(x_t \mid x_{t-1}) =$

	S	H
S	.9	.1
H	0	1

✓ $P(y_t \mid x_t) =$

	W	S	C	F
S	0.3	0.3	0.2	0.2
H	1	0	0	0

$\forall t$



$P(y_1 \mid x_1) = P(y_2 \mid x_2)$

for 2 time steps

$$P(x_0, x_1, x_2, y_1, y_2) = P(x_0)\,P(y_1 \mid x_1)\,P(y_2 \mid x_2)\,P(x_1 \mid x_0)\,P(x_2 \mid x_1)$$

Optimal Filtering

$X_{1:t} = \{X_1, X_2, \dots, X_t\}$

Our goal is to compute, ***for all t****, the posterior (aka* ***filtering****) distribution:*

$$P(X_t \mid Y_{1:t}) = P(X_t \mid \underbrace{Y_1, Y_2, \dots, Y_t}_{})$$

We derive a ***recursion*** *to compute* $P(X_t \mid Y_{1:t})$ *assuming that we have as input* $P(X_{t-1} \mid Y_{1:t-1})$ *. The recursion has two steps:* ***prediction*** *and* ***Bayesian update****.*

$P(X_0)$

S	H
0.2	0.8

$Y_1 \rightarrow P(X_1 \mid Y_1)$

0.1	0.9

$Y_2 \rightarrow P(X_2 \mid Y_1, Y_2) = P(X_2 \mid Y_{1:2})$

1	0

$Y_3 \rightarrow P(X_3 \mid Y_{1:3})$

0.5	0.5

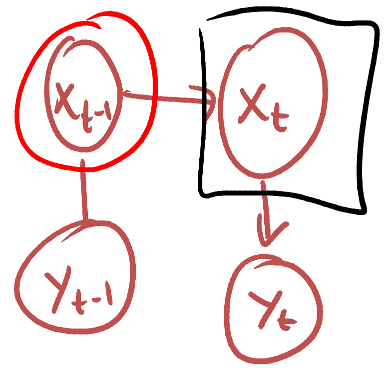
$\vdots$

Prediction

$$P(A|C) = \sum_B P(AB|C)$$

First, we compute the state prediction: **$P(X_t | Y_{1:t-1})$**

$$P(AB|C) = P(A|BC)\,P(B|C)$$



$$P(x_t | y_{1:t-1}) = \sum_{x_{t-1} \in \{H,S\}} P(x_t, x_{t-1} | y_{1:t-1})$$

$$= \sum_{x_{t-1}} P(x_t | x_{t-1}, \cancel{y_{1:t-1}})\, P(x_{t-1} | y_{1:t-1})$$

$$= \sum_{x_{t-1}} P(x_t | x_{t-1})\, P(x_{t-1} | y_{1:t-1})$$

Bayes rule revision

$$P(A|BC) = P(A|CB)$$

$$= \frac{P(ABC)}{P(BC)} = \frac{P(B|AC)P(A|C)P(C)}{P(B|C)\,P(C)}$$

$$= \frac{P(B|AC)P(A|C)}{P(B|C)}$$

Bayes update

Second, we use Bayes rule to obtain $\boldsymbol{P(X_t | Y_{1:t})}$

$$P(x_t \mid y_{1:t}) = P(\overset{A}{x_t} \mid \overset{B}{y_t}, \overset{C}{y_{1:t-1}})$$

$$= \frac{P(y_t \mid x_t, y_{1:t-1})\, P(x_t \mid y_{1:t-1})}{\sum_{x_t} P(y_t \mid x_t, y_{1:t-1})\, P(x_t \mid y_{1:t-1})}$$

$$= \frac{P(y_t \mid x_t)\, P(x_t \mid y_{1:t-1})}{\sum_{x_t} P(y_t \mid x_t)\, P(x_t \mid y_{1:t-1})}$$

HMM algorithm

$\boxed{\text{Example.}}$

Assume $P(X_{t-1}|Y_{1:t-1}) = \begin{array}{cc} \multicolumn{2}{c}{X_{t-1}} \\ S & H \\ \hline .2 & .8 \end{array}$

$P(X_t | X_{t-1}) = \begin{array}{c|cc} & \multicolumn{2}{c}{X_t} \\ & S & H \\ \hline S & 1 & 0 \\ X_{t-1}\ H & 0.1 & 0.9 \end{array}$

$P(Y_t|X_t) = \begin{array}{c|cccc} & \multicolumn{4}{c}{Y_t} \\ & W & S & C & F \\ \hline S & .1 & .2 & .3 & .4 \\ X_t\ H & 0.9 & 0 & 0 & 0.1 \end{array}$

Then, we can predict the state X_t (at time t) given the observations $Y_1, Y_2, \cdots Y_{t-1}$. Specifically

$$P(X_t = H | Y_{1:t-1}) = \sum_{X_{t-1}} P(X_t = H | X_{t-1}) P(X_{t-1} | Y_{1:t-1})$$

$$= P(X_t = H | X_{t-1} = S) P(X_{t-1} = S | Y_{1:t-1}) + P(X_t = H | X_{t-1} = H) P(X_{t-1} = H | Y_{1:t-1})$$

$$= (0)(0.2) + (0.9)(0.8) = 0.72$$

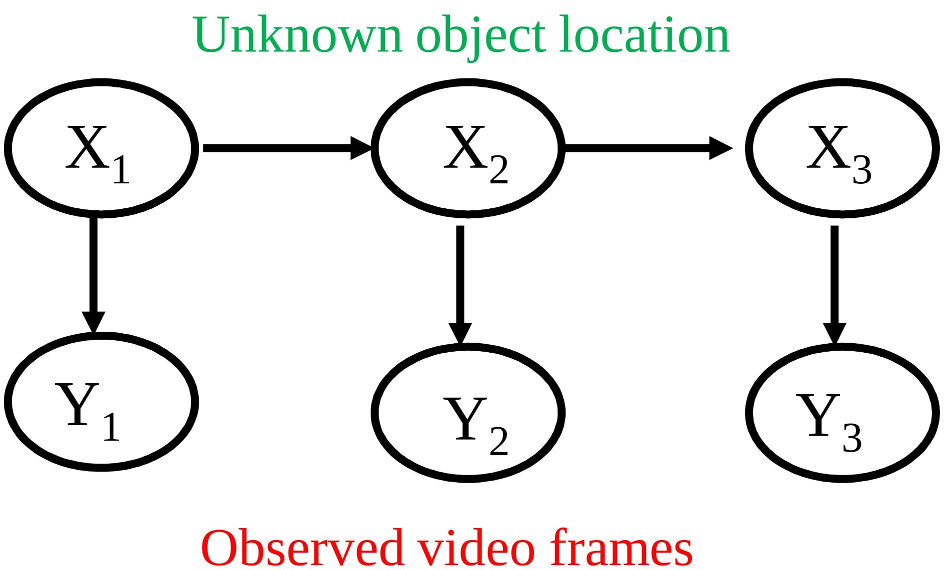
$$P(X_t = S | Y_{1:t-1}) = 1 - P(X_t = H | Y_{1:t-1}) = 0.28$$

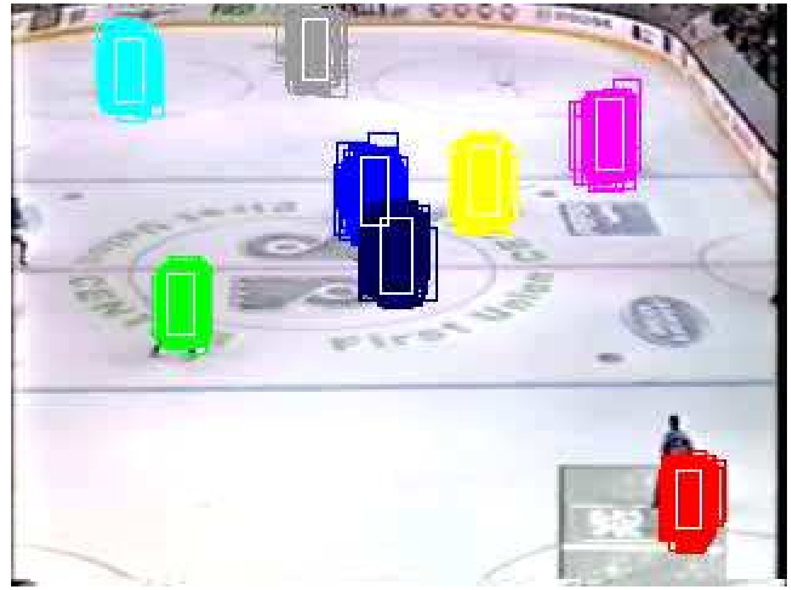
If we observe $Y_t = W$, The Bayes update yields:

$$P(X_t = H | Y_{1:t-1}, Y_t = W) = \frac{P(Y_t = W | X_t = H) P(X_t = H | Y_{1:t-1})}{P(Y_t = W | X_t = H) P(X_t = H | Y_{1:t-1}) + P(Y_t = W | X_t = S) P(X_t = S | Y_{1:t-1})}$$

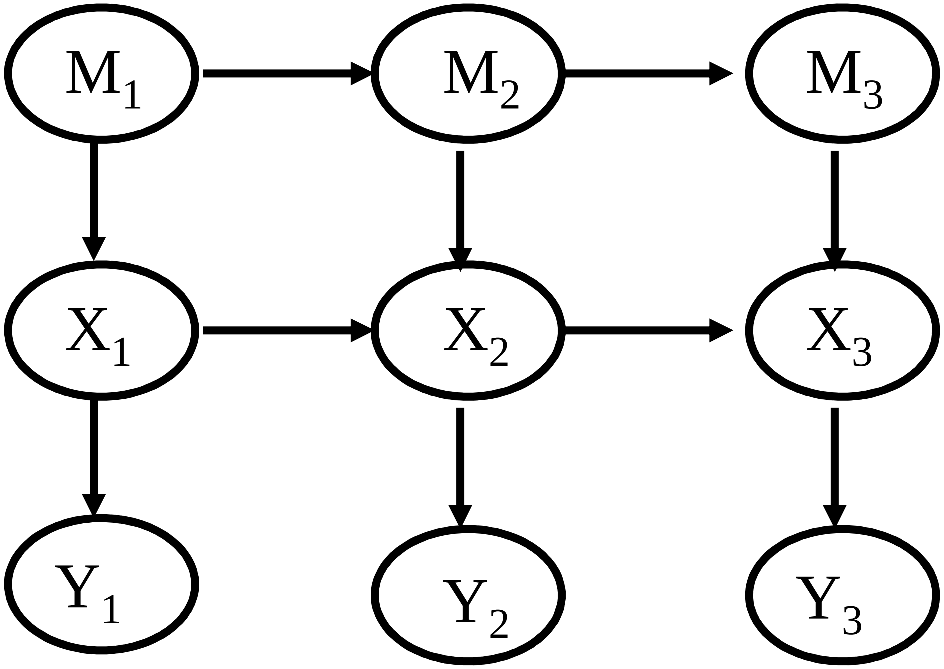
$$= (0.9)(0.72) / [(0.9)(0.72) + (0.1)(0.28)]$$

Example 1: Image tracking





Example 2: Diagnosis

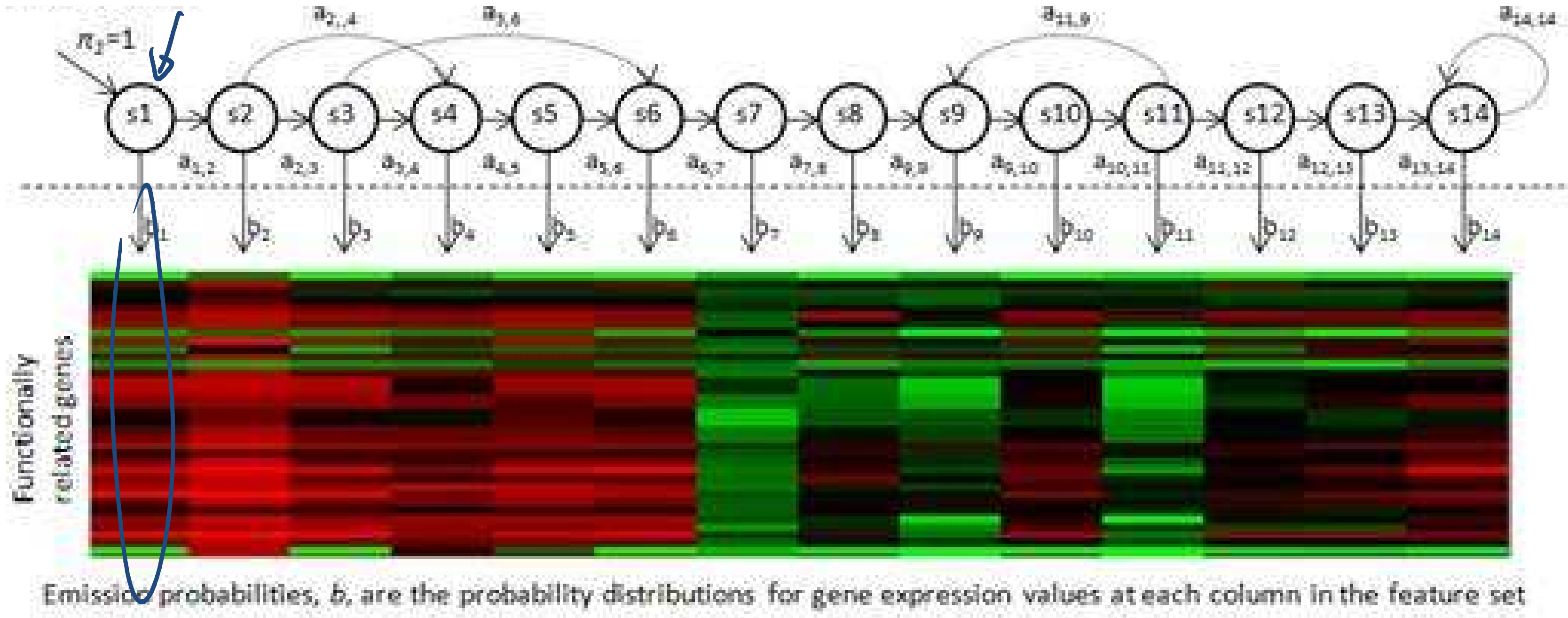


Unknown internal discrete state

Unknown continuous signals

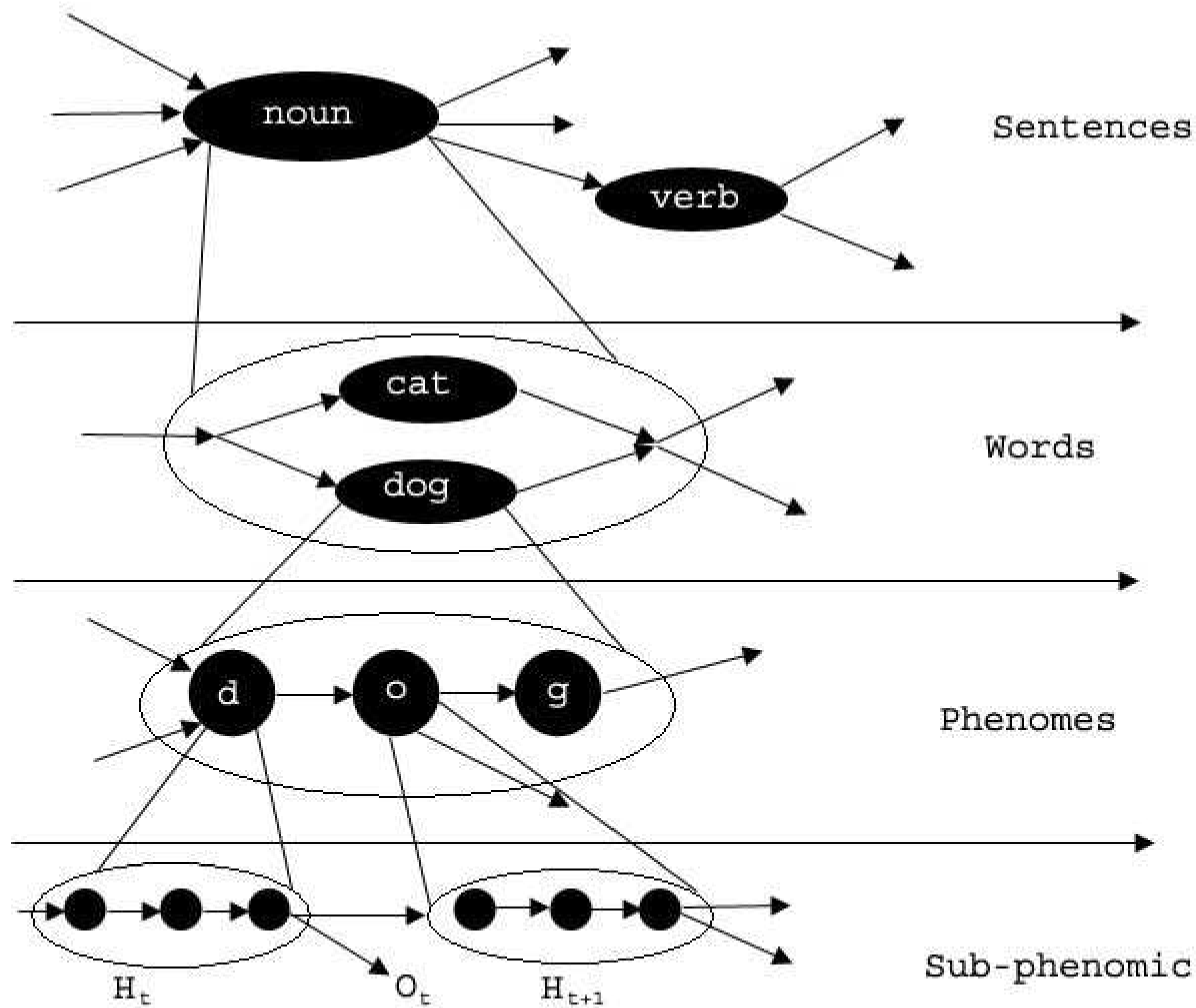
Sensor readings

Example 3: bioinformatics



http://www.ittc.ku.edu/~xwchen/

Example 4: speech recognition



Next lecture

In the next lecture, we revise material needed to attack the problem of learning graphical models from data.