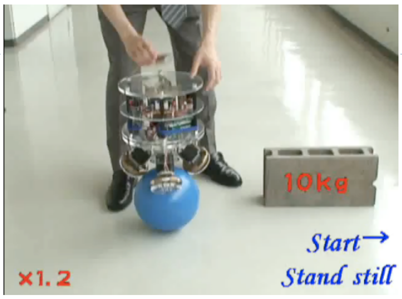


http://www.youtube.com/watch?v=bl06lujiD7E

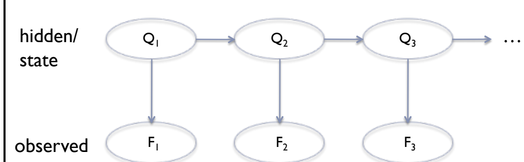
Hidden Markov Models 2

David Kauchak, CS151, Fall 2010

Admin

- Written 6 is out
- Final project literature review, due this Friday
- Final project
  - Project proposal due 11/5
  - If you don't have a partner, send me an e-mail ASAP along with a list of a few topics you might be interested in (or if you don't care)
- Last assignment (HMM) also out
- NB classifier
  - How many people did extra credit?
  - Best performance on data sets/%improvement
  - Interesting/successful features?

Hidden Markov Models



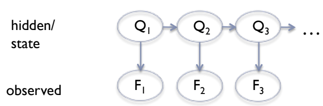
$P(Q_2|Q_1) = P(Q_3|Q_2) = P(Q_4|Q_3) = P(Q_t|Q_{t-1})$

$P(F_1|Q_1) = P(F_2|Q_2) = P(F_3|Q_3) = P(F_t|Q_t)$

Stationary process: value of the state may change but not the underlying distribution

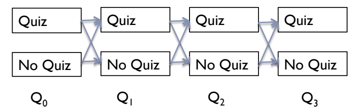
What questions could we ask?

- Filtering: What is the distribution over Q today?
  - $P(Q_t \mid F_{1:t})$
- Prediction: What is the distribution Q in the future?
  - $P(Q_{t+k} \mid F_{1:t})$ – k days from now
- Smoothing: What is the distribution over Q in the past?
  - $P(Q_k \mid F_{1:t})$ – for k < t
- Most likely explanation: What states have we been in?
  - $\text{argmax}_{Q1:t}\, P(Q_{1:t} \mid F_{1:t})$
- Learning
  - Learn CPD tables given observations



Trellis

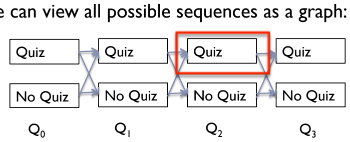
We can view all possible sequences as a graph:



The "trellis" describes all the possible path

Filtering revisited

We can view all possible sequences as a graph:



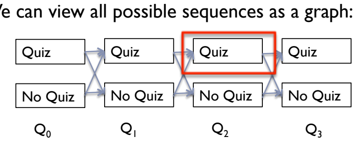
The "trellis" describes all the possible path

Does the trellis help explain the filtering equation?

Filtering: $P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$

Filtering revisited

We can view all possible sequences as a graph:



Filtering: $P(q_2 \mid F_{1:2}) = \alpha P(f_2 \mid q_2) \sum_{q_1} P(q_2 \mid q_1) P(q_1 \mid F_1)$

- There are two ways we could get to having a quiz on day 2, either quiz or no quiz on day 1.
- We don't know which of these happened, so we sum over them, multiplying by the probability of them happening!
- We then multiply times the probability of a quiz on the second day give our observed data.

Smoothing

- Smoothing: What is the distribution over Q in the past?
  - P(Q_k | $F_{1:t}$) – for k < t

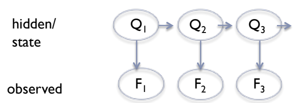
Smoothing

- Smoothing: What is the distribution over Q in the past?
  - P(Q_k | $F_{1:t}$) – for k < t

$$P(Q_k \mid F_{1:t})$$

$$\alpha P(Q_k \mid F_{1:k}) P(F_{k+1:t} \mid Q_k)$$

Try and derive this second equation. Tools:
- Bayes Rule
- Split out variables ($F_{1:t} = F_{1:t-1}, F_t$)
- Bayes Net structure (conditional independence)
- introduce and sum over a new variable



Smoothing

- Smoothing: What is the distribution over Q in the past?
  - P(Q_k | $F_{1:t}$) – for k < t

$$\begin{aligned} P(Q_k \mid F_{1:t}) &= P(Q_k \mid F_{1:k}, F_{k+1:t}) \\ &= \alpha P(Q_k \mid F_{1:k}) P(F_{k+1:t} \mid Q_k, F_{1:k}) && \text{Bayes' rule} \\ &= \alpha \underbrace{P(Q_k \mid F_{1:k})}_{\text{forward message (Filtering)}} \underbrace{P(F_{k+1:t} \mid Q_k)}_{\text{backward message}} && \text{BN structure} \end{aligned}$$

Smoothing

- Smoothing: What is the distribution over Q in the past?
  - P(Q_k | $F_{1:t}$) – for k < t

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) P(F_{k+1:t} \mid Q_k)$$

$$P(F_{k+1:t} \mid Q_k)$$

$$\sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

Try and derive this second equation. Tools:
- Bayes Rule
- Split out variables ($F_{1:t} = F_{1:t-1}, F_t$)
- Bayes Net structure (conditional independence)
- introduce and sum over a new variable

Smoothing

- Smoothing: What is the distribution over Q in the past?
  - $P(Q_k \mid F_{1:t})$ – for k < t

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) P(F_{k+1:t} \mid Q_k)$$

$$P(F_{k+1:t} \mid Q_k) = \sum_{q_{k+1}} P(F_{k+1:t} \mid Q_k, q_{k+1}) P(q_{k+1} \mid Q_k)$$

$$= \sum_{q_{k+1}} P(F_{k+1:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$ BN structure

$$= \sum_{q_{k+1}} P(F_{k+1}, F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$ splitting up the evidence

$$= \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$ BN structure

Smoothing

- Smoothing: What is the distribution over Q in the past?
  - $P(Q_k \mid F_{1:t})$ – for k < t

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) P(F_{k+1:t} \mid Q_k)$$

What is this?

$$P(F_{k+1:t} \mid Q_k) = \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

probability of folder given quiz

probability of quiz given quiz yesterday

Can be read from our CPTs

Smoothing

- Smoothing: What is the distribution over Q in the past?
  - $P(Q_k \mid F_{1:t})$ – for k < t

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) P(F_{k+1:t} \mid Q_k)$$

$$P(F_{k+1:t} \mid Q_k) = \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

Recursive case!
(message)

Quiz taking

- What is the probability of a quiz today given no folder today or last class?
  - $P(Q_2 \mid NF, NF)$ = [0.165, 0.835]
    Quiz No Quiz
- On the next class, I bring in the dreaded folder. What now is the probability that we had a quiz that previous class?
  - $P(Q_2 \mid NF, NF, F)$ = ?

Smoothing

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

Smoothing

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

$$= \alpha P(Q_2 \mid \neg f, \neg f)\left(P(f \mid q) P(F_{k+2:t} \mid q) P(q \mid Q_k) + P(f \mid \neg q) P(F_{k+2:t} \mid \neg q) P(\neg q \mid Q_k)\right)$$

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

$$= \alpha \begin{bmatrix} 0.165 \\ 0.835 \end{bmatrix} \left(0.8 \cdot 1 \cdot \begin{bmatrix} 0.1 \\ 0.5 \end{bmatrix} + 0.3 \cdot 1 \cdot \begin{bmatrix} 0.9 \\ 0.5 \end{bmatrix} \right)$$

$$= \alpha \begin{bmatrix} 0.0578 \\ 0.459 \end{bmatrix} = \begin{bmatrix} 0.112 \\ 0.888 \end{bmatrix}$$

Quiz taking

- $P(Q_2 \mid NF, NF) = [0.165, 0.835]$
  Quiz No Quiz

Do these probabilities make sense?

- $P(Q_2 \mid NF, NF, F) = [0.112, 0.888]$
  Quiz No Quiz

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

Quiz taking

- $P(Q_2 \mid NF, NF) = [0.165, 0.835]$
  Quiz No Quiz
- $P(Q_2 \mid NF, NF, F) = [0.112, 0.888]$
  Quiz No Quiz

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

The likelihood of a quiz, given that I brought the folder is high.

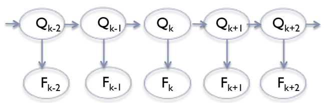
The likelihood of a two quizzes in a row, is low.

Therefore, it becomes less likely that I had a quiz on day 2.

Smoothing as recursive messages

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

$P(Q_k|F_{1:t})$

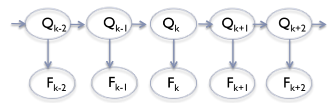


Calculating the $P(Q_k \mid F_{1:t})$ involves recursion both forward and backwards

Smoothing as recursive messages

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

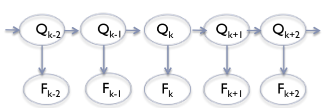
$P(Q_{k-1}|F_{1:k-1})$ $P(Q_k|F_{1:t})$ $P(F_{k+2:t}|Q_{k+1})$



Smoothing as recursive messages

$$P(Q_k \mid F_{1:t}) = \alpha P(Q_k \mid F_{1:k}) \sum_{q_{k+1}} P(F_{k+1} \mid q_{k+1}) P(F_{k+2:t} \mid q_{k+1}) P(q_{k+1} \mid Q_k)$$

$P(Q_{k-2}|F_{1:k-2})$ $P(Q_{k-1}|F_{1:k-1})$ $P(Q_k|F_{1:t})$ $P(F_{k+2:t}|Q_{k+1})$ $P(F_{k+3:t}|Q_{k+2})$



What if we want to smooth all states, 1 through t, that is calculate $P(Q_k \mid F_{1:t})$ for all k?

Forward-Backward algorithm

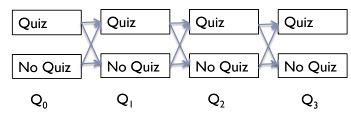
- Option 1: Do the previous approach for each 1…t. Any problems?
  - inefficient! We end up recalculating a lot of things, in particular the forward and backward messages
- Option 2: Forward-Backward algorithm
  - Start at time 1 and calculate all of the forward messages going forward and save them
  - Then, from t backward
    - calculate the backward message
    - calculate the smoothed probability at that node (we now have both the forward and backward messages

Most likely explanation

- Most likely explanation: What states have we been in?
  - $\text{argmax}_{Q_{1:t}} P(Q_{1:t} \mid F_{1:t})$
- Ideas?
  - Option 1:
    - $\text{argmax}_{Q_{1:t}} P(Q_{1:t} \mid F_{1:t}) = \text{argmax}_{Q_{1:t}} P(Q_1 \mid F_{1:t}) P(Q_2 \mid F_{1:t}) \ldots P(Q_t \mid F_{1:t})$
    - calculate the smoothed value $P(Q_k \mid F_{1:t})$ for all k. Pick the argmax at each step
  - problems?
    - Locally greedy. Doesn't take into account the actual sequence chosen, specifically $P(X_t \mid X_{t-1})$

Trellis

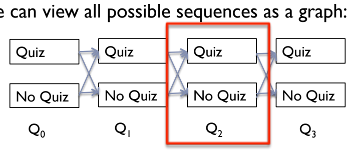
We can view all possible sequences as a graph:



We'd like to find the most probable path

Most likely explanation

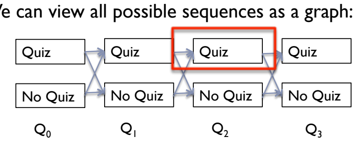
We can view all possible sequences as a graph:



Consider any day. As with filtering, for each option, we just need to consider the previous day (Markov assumption)

Most likely explanation

We can view all possible sequences as a graph:



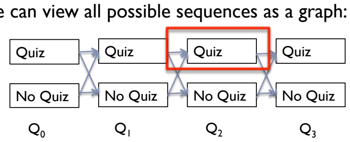
Filtering: $P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$

Filtering summed over all possible previous states

We want the *most likely*. What should we do?

Most likely explanation

We can view all possible sequences as a graph:



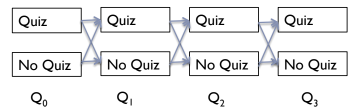
We have two choices for today, Quiz or No Quiz. Focus on just one, say Quiz.

Because of the Markov assumption, the most likely path with a Quiz today came from either Quiz or No Quiz yesterday.

Check both of these options and **pick** the *larger*

Most likely explanation

We can view all possible sequences as a graph:



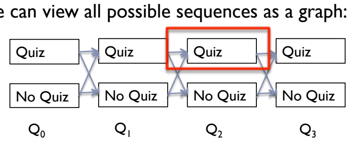
Filtering: $P(Q_t \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \sum_{q_{t-1}} P(Q_t \mid q_{t-1}) P(q_{t-1} \mid F_{1:t-1})$

current observation — state transition — recursive case

$$\max_{Q_{1:t}} P(Q_{1:t} \mid F_{1:t}) = \alpha P(F_t \mid Q_t)\left(\max_{q_{t-1}} P(Q_t \mid q_{t-1}) \max_{Q_{1:t-1}} P(Q_{1:t-1} \mid F_{1:t-1})\right)$$

Most likely explanation

We can view all possible sequences as a graph:



$$\max_{Q_{1:t}} P(Q_{1:t} \mid F_{1:t}) = \alpha P(F_t \mid Q_t)\left(\max_{q_{t-1}} P(Q_t \mid q_{t-1}) \max_{Q_{1:t-1}} P(Q_{1:t-1} \mid F_{1:t-1})\right)$$

$$P(q_2 \mid F_{1:2}) = \alpha P(f_2 \mid q_2) \underbrace{\max_{q_1} P(q_2 \mid q_1)} \underbrace{\max_{Q_{1:1}} P(Q_{1:1} \mid f_1)}$$

Pick the most likely sequence for this state: q, q or q, nq

Recursive case: gives us the probabilities of the most likely sequences ending in the previous state

Most likely explanation

- Similar to filtering
  - Start at the beginning and move to end
    - Calculate the "message" ($P(q_{1:i} \mid F_{1:i})$) given previous message
    - Note that this is a set of probabilities, one for each state value
  - If you want the actual choice, need to store backpointers
- Called the Viterbi algorithm

$$\max_{Q_{1:t}} P(Q_{1:t} \mid F_{1:t}) = \alpha \underbrace{P(F_t \mid Q_t) \max_{q_{t-1}} P(Q_t \mid q_{t-1})} \underbrace{\max_{Q_{1:t-1}} P(Q_{1:t-1} \mid F_{1:t-1})}$$

Notice that this gives us a distribution over Q_t

Recursive case: gives us the probabilities of the most likely sequences ending in the previous state

Did we have a quiz?

- The student has walked by the window for the past three days and has seen no folder, no folder and folder. Have we had a quiz? Assume a uniform prior.

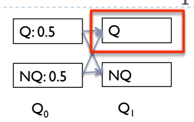
Did we have a quiz on the first day?

Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

$$\max_{Q_{1:t}} P(Q_{1:t} \mid F_{1:t}) = \alpha P(F_t \mid Q_t) \max_{q_{t-1}} P(Q_t \mid q_{t-1}) \max_{Q_{1:t-1}} P(Q_{1:t-1} \mid F_{1:t-1})$$

Did we have a quiz on the first day?



Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

message

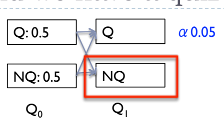
$$P(q_1 \mid F_1) = \alpha P(nf \mid q_1) \max_{q_0} P(q_1 \mid q_0) \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$= \alpha\, 0.2 \ \max_{q_0} \{P(q_1 \mid q_0) 0.5,\ P(q_1 \mid \neg q_o) 0.5\}$$

$$= \alpha\, 0.2 \ \max_{q_0} \{0.1 \cdot 0.5,\ 0.5 \cdot 0.5\}$$

$$= \alpha\, 0.2\ 0.25 = 0.05$$

Did we have a quiz on the first day?



Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

message

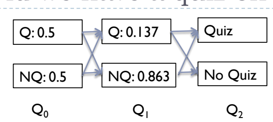
$$P(\neg q_1 \mid F_1) = \alpha P(nf \mid \neg q_1) \max_{q_0} P(\neg q_1 \mid q_0) \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

$$= \alpha\, 0.7 \ \max_{q_0} \{P(\neg q_1 \mid q_0) 0.5,\ P(\neg q_1 \mid \neg q_o) 0.5\}$$

$$= \alpha\, 0.7 \ \max_{q_0} \{0.9 \cdot 0.5,\ 0.5 \cdot 0.5\}$$

$$= \alpha\, 0.7\ 0.45 = \alpha\, 0.315$$

Did we have a quiz on the first day?



Q_{t-1}	Q_t
T	0.1
F	0.5

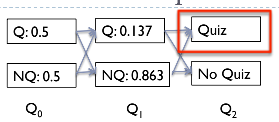
Q_t	F_t
T	0.8
F	0.3

Did we have a quiz on the first day?

If we didn't have any other information, most likely, not.

However, we don't know yet with additional information!

Did we have a quiz on the 2nd day?



Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

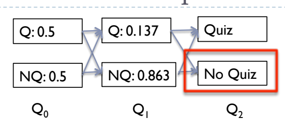
$$P(q_2 \mid F_1) = \alpha P(nf \mid q_2)\max_{q_1} P(q_2 \mid q_1) message(Q_1)$$

$$= \alpha\, 0.2\ \max_{q_0}\{P(q_2 \mid q_1)0.137,\ P(q_2 \mid \neg q_1)0.863\}$$

$$= \alpha\, 0.2\ \max_{q_0}\{0.1 \cdot 0.137,\ 0.5 \cdot 0.863\}$$

$$= \alpha\, 0.2\ 0.431 = 0.0862$$

Did we have a quiz on the 2nd day?



Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

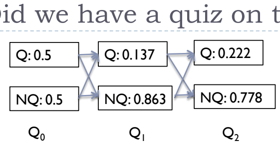
$$P(\neg q_2 \mid F_1) = \alpha P(nf \mid \neg q_2)\max_{q_1} P(\neg q_2 \mid q_1) message(Q_1)$$

$$= \alpha\, 0.7\ \max_{q_0}\{P(\neg q_2 \mid q_1)0.137,\ P(\neg q_2 \mid \neg q_1)0.863\}$$

$$= \alpha\, 0.7\ \max_{q_0}\{0.9 \cdot 0.137,\ 0.5 \cdot 0.863\}$$

$$= \alpha\, 0.7\ 0.431 = 0.302$$

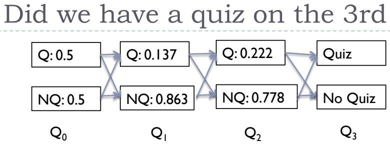
Did we have a quiz on the 2nd day?



Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

Did we have a quiz on the 3rd day?



Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

$$P(q_3 \mid nf, nf, f) = \alpha P(f \mid q_3)\ \max_{q_2}\{P(q_3 \mid q_2)0.222,\ P(q_2 \mid \neg q_1)0.778\}$$

$$= \alpha\, 0.8\ \max_{q_2}\{0.1 \cdot 0.222,\ 0.5 \cdot 0.778\}$$

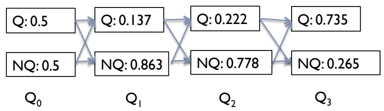
$$= \alpha\, 0.311$$

$$P(\neg q_2 \mid nf, nf, f) = \alpha P(f \mid \neg q_2)\max_{q_2}\{P(\neg q_3 \mid q_2)0.222,\ P(\neg q_3 \mid \neg q_2)0.778\}$$

$$= \alpha\, 0.3\ \max_{q_0}\{0.9 \cdot 0.357,\ 0.5 \cdot 0.743\}$$

$$= \alpha\, 0.112$$

Did we have a quiz on the 3rd day?

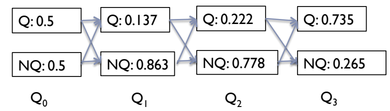


Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

What is the most likely sequence?

Did we have a quiz on the 3rd day?



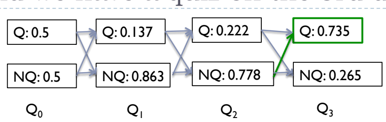
Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

Be careful! All that we have calculated is the probability of the most likely sequence.

If we want the sequence, we need to store back pointers (dynamic programming) as we go.

Did we have a quiz on the 3rd day?



Q_{t-1}	Q_t
T	0.1
F	0.5

Q_t	F_t
T	0.8
F	0.3

$$P(\neg q_2 \mid nf, nf, f) = \alpha P(f \mid \neg q_2) \max_{q_2} \{P(\neg q_3 \mid q_2) 0.222,\ P(\neg q_3 \mid \neg q_2) 0.778\}$$

$$= \alpha\, 0.3 \max_{q_0} \{0.9 \cdot 0.357,\ 0.5 \cdot 0.743\}$$

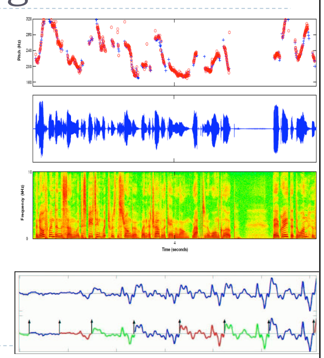
$$= \alpha\, 0.112$$

In our recursive max, we chose one of the values for the previous state

Speech Recognition

HMM problem?

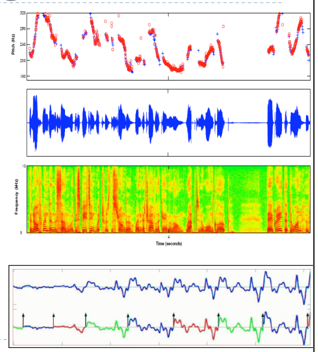
- What is observed?
- What are the states?
- What is the key HMM question?



S. Roweis, 2004

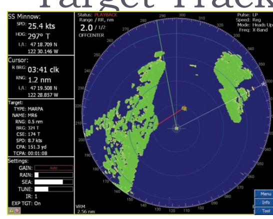
Speech Recognition

- States
  - words (i.e. all of the possible English words)
  - phonemes
- Observed
  - The wave form, specifically, extract features over some time interval
- Question:
  - most likely sequence
- HMMs do well, but not state of the art

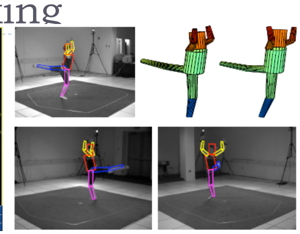


S. Roweis, 2004

Target Tracking



Radar-based tracking of multiple targets

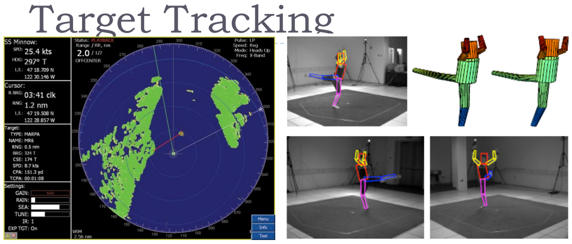


Visual tracking of articulated objects
(L. Sigal et. al., 2006)

HMM problem?

What is observed?
What are the states?
What is the key HMM question?

Target Tracking



- Estimate motion of targets in 3D world from indirect, potentially noisy measurements
- Observed: noisy information (radar, from visual data)
- States: actual position of plane/person/arm
- Questions: smoothing, most likely sequence

Part of speech tagging

- Parts of speech:
  - Nouns: people, animals, concepts, things
  - Verbs: express action in the sentence
  - Adjectives: describe properties of nouns

Input: The lead paint is unsafe

Output: The/DET lead/N paint/N is/V unsafe/ADJ

HMM problem?

What is observed?
What are the states?
What is the key HMM question?

Part of speech tagging

- Parts of speech:
  - Nouns: people, animals, concepts, things
  - Verbs: express action in the sentence
  - Adjectives: describe properties of nouns

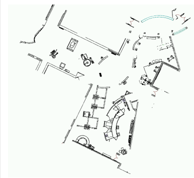
Input: The lead paint is unsafe

Output: The/DET lead/N paint/N is/V unsafe/ADJ

  - Observed
    - words
  - States
    - parts of speech
  - HMM task
    - most likely sequence

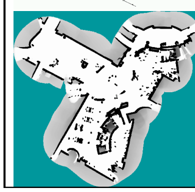
Robot Navigation: *SLAM*

Simultaneous Localization and Mapping

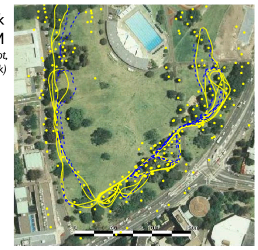


CAD Map

(S.Thrun, San Jose Tech Museum)



Estimated Map

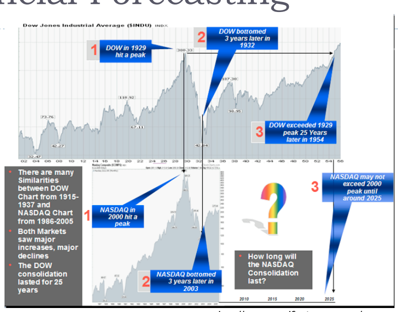


Landmark SLAM

(E. Nebot, Victoria Park)

- As robot moves, estimate its pose & world geometry

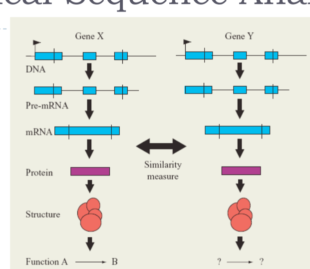
Financial Forecasting



http://www.steadfastinvestor.com/

- Predict future market behavior from historical data, news reports, expert opinions, …

Biological Sequence Analysis

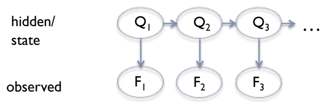


(E. Birney, 2001)

- Temporal models can be adapted to exploit more general forms of *sequential* structure, like those arising in DNA sequences

More complicated Markov models

- Is the first-order Markov assumption appropriate for our quiz example?



Probably not. It's likely that we'd need to incorporate more than just the previous day's state.

Any way to fit this within our HMM framework?

More complicated Markov models

- We can redefine what the state is
- Say we would like to have $P(Q_t \mid Q_{t-1}, Q_{t-2})$ (that is a second-order Markov model)
  - Rather than just having the state be Q_t, we let the state be $Q_t \times Q_{t-1}$
  - Then, our CPT would be $P(Q_t \times Q_{t-1} \mid Q_{t-1} \times Q_{t-2})$ which you can use as just $P(Q_t \mid Q_{t-1} \times Q_{t-2})$

More complicated Markov Models

- Often, we will have more than just one feature that is observed
- Quiz
  - whether I had a quiz in my other class
  - whether I look stressed or not
  - the week in the semester
  - if we just had a midterm (if we have an upcoming midterm)
  - …
- The CPD then becomes $P(feature_1, feature_2, \ldots, feature_n \mid Qt)$
- An HMM, assumes a CPD with all possible feature combination
- A dynamic bayes net, can simplify this assumption
  - in our lab, we assume features are independent given the state