10/6/16

HMM & the Forward Algo

Brendan O'Connor UMass Amherst

CS 585

How to build a POS tagger?

other sequence tagging tasks

I saw B. O. go to the store

NotPER PER PER

Named Entity Recog.

- Key sources of information:
  - HMM
  - 1. The word itself — to chair ~~the~~ session
    Prep N .. V ? ??
  - 2. Word-internal characters → CRF
    chaired "word ends with -ed"
  - 3. POS tags of surrounding words: syntactic context
    ... attack
    Det Noun

Markov Model

$w \to w \to w$ $\quad P(\vec{w}) = \prod_{t=1}^{T} P(w_t | w_{t-1})$

Hidden Markov Model

$y_1 \dots y_T$: hidden/latent states

$$\begin{tikzcd} y_1 \arrow[r] \arrow[d] & y_2 \arrow[r] \arrow[d] & \dots \\ w_1 & w_2 & \end{tikzcd}$$

$P(\vec{w}, \vec{y})$

observed words $\quad$ hidden states

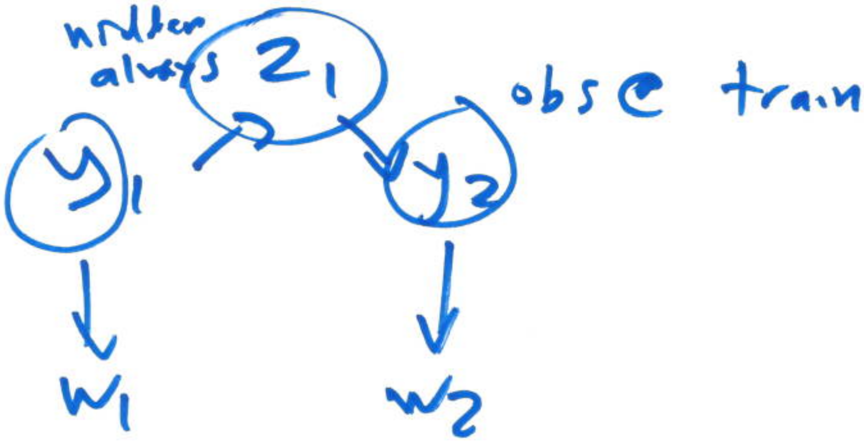
$$P(\vec{w}, \vec{y}) = \prod_{t=1}^{T} \underbrace{P(y_t | y_{t-1})}_{\text{Transition Model}} \underbrace{P(w_t | y_t)}_{\text{Emission/Observation Model}}$$

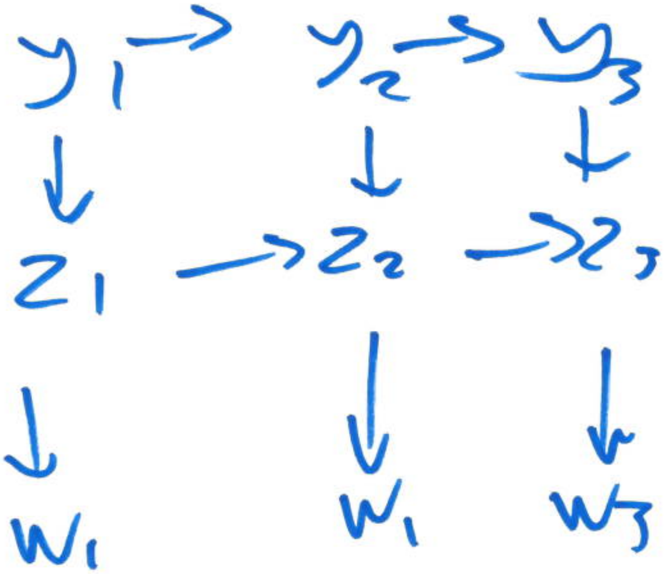
other Examples?

Econ: y = Recession? $\quad$ w = Employment Stats

Eco: y = ~~Animal~~ Bird Pop. $\quad$ w = Reports of Sightings

y = Temp/weather $\quad$ w = Clothing Types





"Factoral HMM"

What good is an HMM?

Model: $P(\vec{w}, \vec{y}) = \prod_t \underbrace{P(y_t \mid y_{t-1})} \underbrace{P(w_t \mid y_t)}$

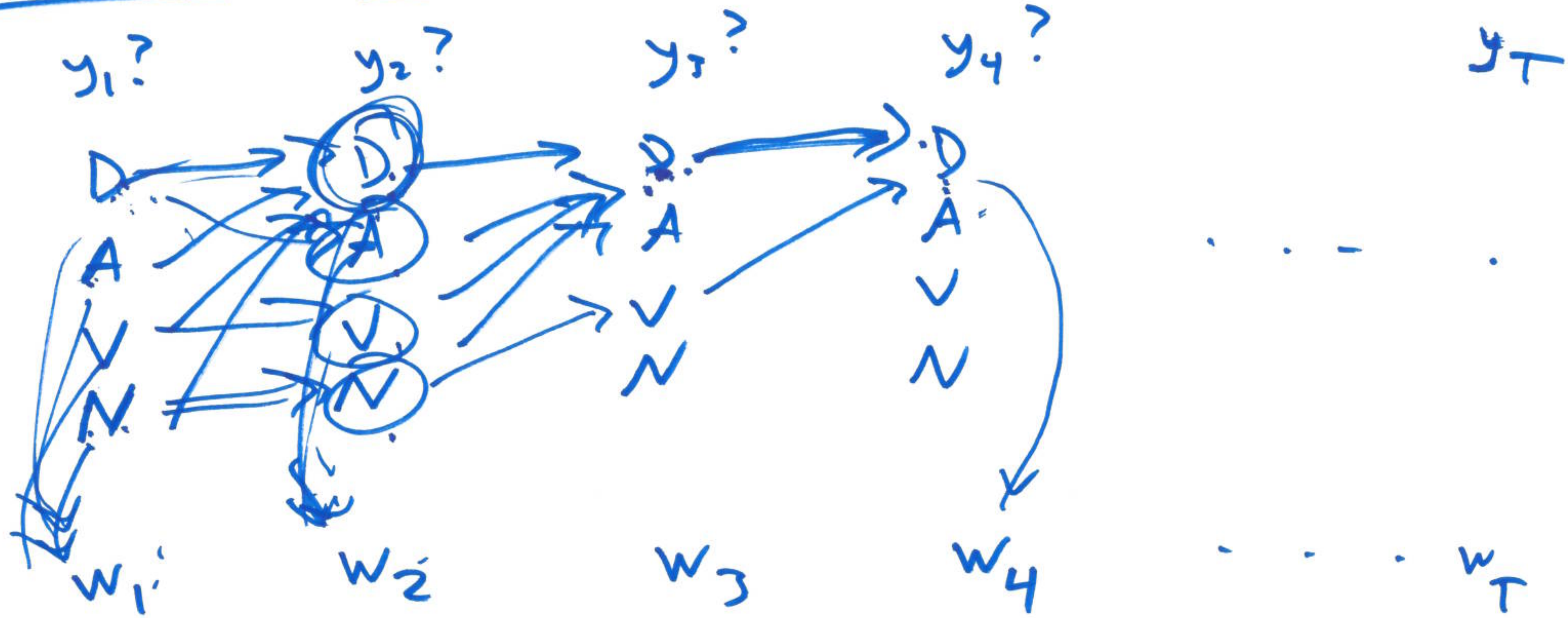
① Likelihood (LM) calc $P(\vec{w}) = \underbrace{\sum_{\vec{y}}}_{\cancel{\text{brg sum}}} \underline{P(\vec{w}, \vec{y})}$

Algo: Forward Algo

② Tagging/Pred/Decoding: calc $\operatorname{argmax}_{\vec{y}} P(\vec{y} / \vec{w})$

Algo: Viterbi

Fwd Algo **Goal:** calc $P(\vec{w}) = \sum_{\vec{y}} P(\vec{y}, \vec{w})$



Exhaustive bad: K^T paths

Idea: Incrementally sum out paths from left to right

$P(\vec{w}) = \sum_{\vec{y}} P(\vec{w}, \vec{y})$

Handout 10/6/16
From J&M -- Jason Eisner's ice cream / weather HMM example.

Model

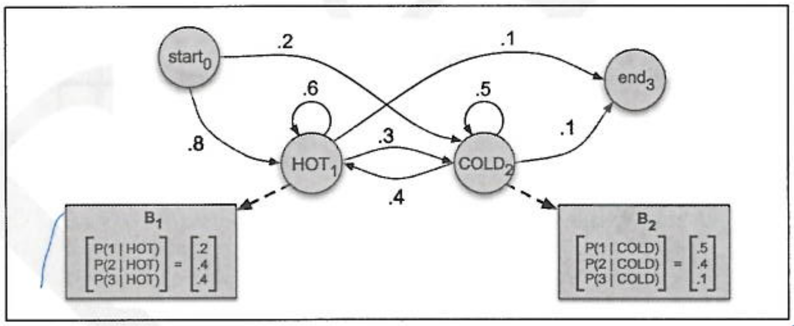


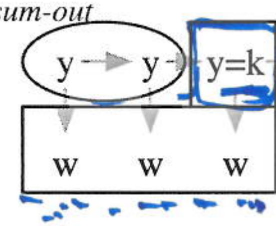
Figure 7.3 A hidden Markov model for relating numbers of ice creams eaten by Jason (the observations) to the weather (H or C, the hidden variables).

K state types
length T seq
Runtime?
$O(T K^2)$

Forward algorithm

Declaratively:

$$\alpha_t[k] = \sum_{y_1 \ldots y_{t-1}} P(y_t = k,\ w_1..w_t,\ y_1..y_{t-1})$$



K^3 for second-order HMM

Recursive Algo.: for each *t=1..N*,

$\forall k$
$$\alpha_t[k] := \sum_{j=1..K} \Big(\alpha_{t-1}[j]\ P_{trans}(k \mid j)\ P_{emit}(w_t \mid k) \Big)$$

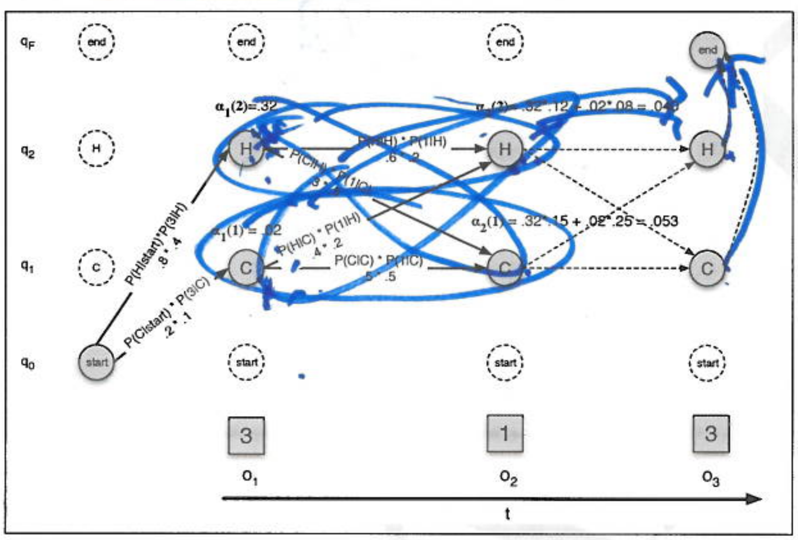
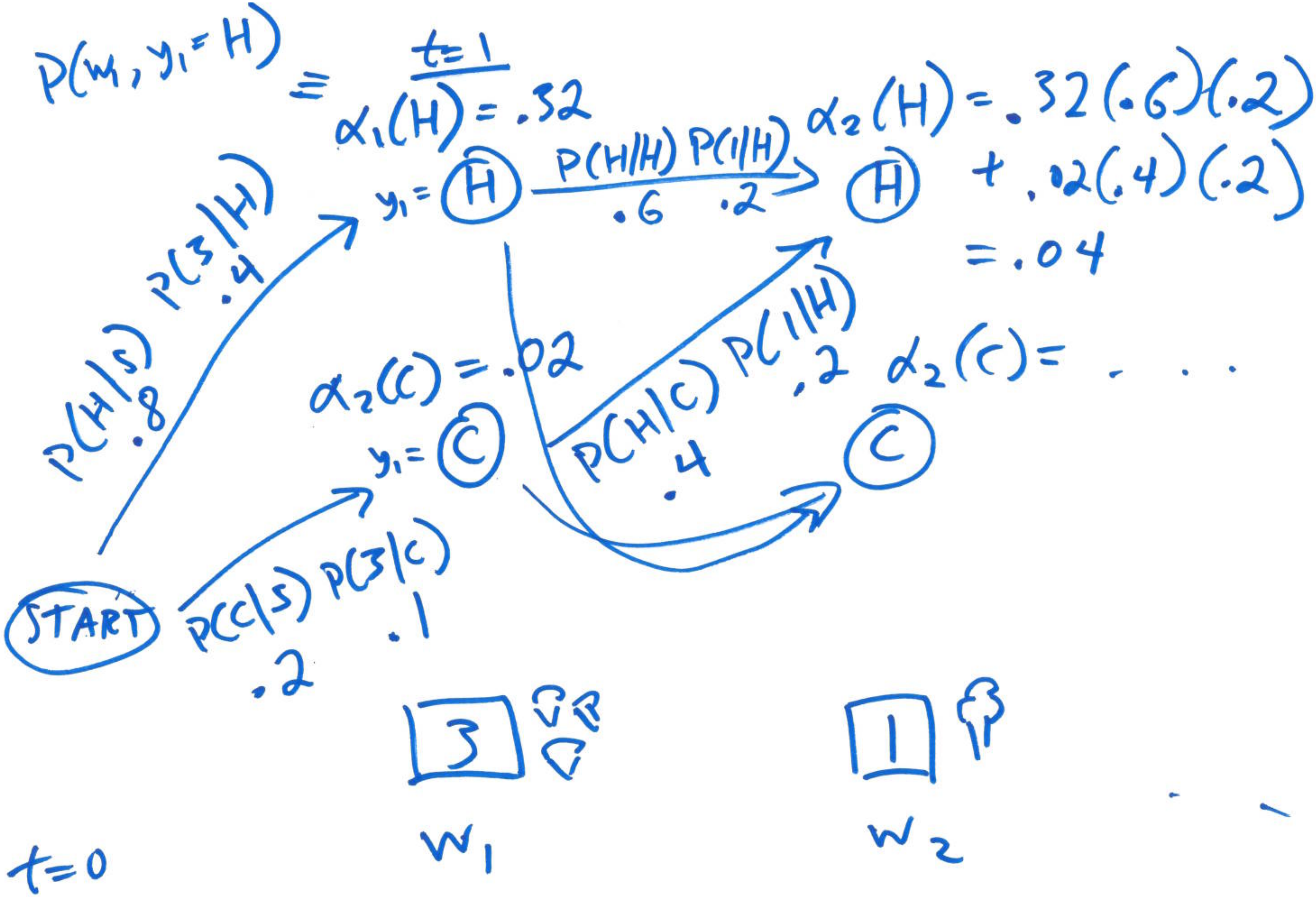


Figure 7.7 The forward trellis for computing the total observation likelihood for the ice-cream events *3 1 3*. Hidden states are in circles, observations in squares. White (unfilled) circles indicate illegal transitions. The figure shows the computation of $\alpha_t(j)$ for two states at two time steps. The computation in each cell follows Eq. 7.14: $\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} b_j(o_t)$. The resulting probability expressed in each cell is Eq. 7.13: $\alpha_t(j) = P(o_1, o_2 \ldots o_t, q_t = j|\lambda)$.



P(w1, y1=H)
≡
t=1
α1(H) = .32
α2(H) = .32(.6)(.2)
+ .02(.4)(.2)
= .04
P(H|H) P(1|H)
.6 .2
y1 = H
P(H|S) .8
P(3|H) .4
α2(C) = .02
y1 = C
P(H|C) .4
P(1|H) .2
α2(C) = . . .
START
P(C|S) .2
P(3|C) .1
3
1
w1
w2
t=0