How to build a POS tagger?

UMass CS585 9/26/17

- Key sources of information:
  - 1. The word itself
  - 2. Word-internal characters
  - 3. POS tags of surrounding words: syntactic context
- Approach: supervised learning (text => tags)
  - Today/Thursday: with the Hidden Markov Model
  - Next week: Conditional Random Field (arbitrary features)

Sequence Tagging

	I	saw	Fred	Smith	go	to	the	store
$\vec{y}$	O	O	B-PER	I-PER	O	O	O	O

- Named Entity Recog with "BIO" tagging
- Part-of-Speech tagging

$\vec{w} \Rightarrow \vec{y}$

input output

Markov Model

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

Prev Word

Hidden Markov Model

$y_1 \to y_2 \to y_3 \to \ldots$

$\downarrow \quad \downarrow \quad \downarrow$

$w_1 \quad w_2 \quad w_3$

$y_1 \ldots y_T$: hidden states

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

$K \times K$

$V \times K$

HMM 9/26/

UMass CS 59

other Examples

- Music
- Economics $\quad y$ = Recession? $\quad w$ = Econ stats
- Radar Tracking $\quad y$ = Ship loc. $\quad w$ = obs. signal

HMM uses $P(\vec{w}, \vec{y}) = \prod_t P(w_t|y_t) P(y_t|y_{t-1})$

Learning: Supervised learning with (w^{tr}, y^{tr})

Inference

① Likelihood (LM): $P(\vec{w}) = \sum_{\vec{y}} P(\vec{w}, \vec{y})$

K^T

$K = \#$ latent states

T = sent. length

Effic. Algo: Forward Algo.

② Tagging / Prediction: $\underset{\vec{y}}{\text{argmax}}\ P(\vec{y}|\vec{w})$

Effic. Algo: Viterbi Algo.