

How to build a POS tagger?

UMass CS588
9/26/17

- Key sources of information:

Tag Dict

- 1. The word itself

Affix-going

gyred

- 2. Word-internal characters

axe

- 3. POS tags of surrounding words:

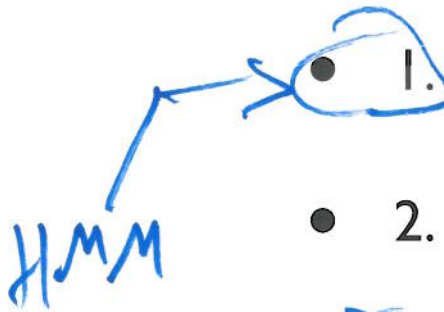
syntactic context

the
same
a
that
Det

attack
N or V

seq tagging

- Approach: supervised learning (text => tags)
 - Today/Thursday: with the Hidden Markov Model
 - Next week: Conditional Random Field (arbitrary features)



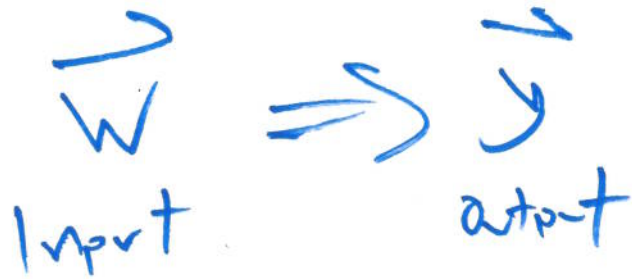
Sequence Tagging

13
12
11
10
9

| | | | | | | | |
|---|-----|-------|-------|----|----|-----|-------|
| I | saw | Fred | Smith | go | to | the | store |
| 0 | 0 | B-PER | I-PER | 0 | 0 | 0 | 0 |

→ Named Entity Recog with "BIO" tagging

- Part-of-Speech tagging



Markov Model

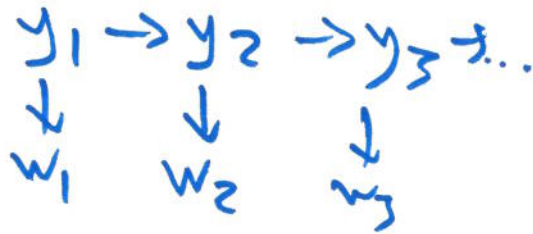
$$w \rightarrow Sw \rightarrow w$$

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

~~etc~~ ↑
Prev Word

Hidden Markov Model

$y_1 \dots y_T$: hidden states



$$P(\vec{w}, \vec{y}) = \prod_t \underbrace{P(y_t | y_{t-1})}_{\substack{\text{Transition Model} \\ K \times K}} \underbrace{P(w_t | y_t)}_{\substack{\text{Observation Model} \\ \text{Emission} \\ V \times K}}$$

HMM
9/26/1
U-Mass
CS 535

Other Examples

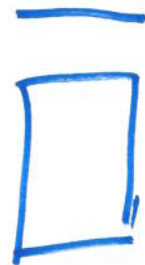
- Music
- Economics
- Radar Tracking

y = Recession?

w = Econ Stats

y = Ship loc.

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

$K \uparrow$
 T

$K = \#$ latent classes
 $T = \text{sent. length}$

Eff.:

Algo: Forward Algo.

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

Eff.:

Algo: Viterbi Algo.