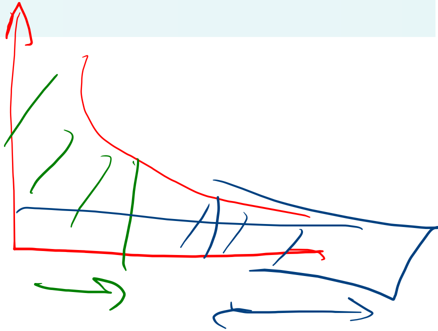
Lecture reviews — Week 07

J.-C. Chappelier & M. Rajman

Laboratoire d'Intelligence Artificielle
Faculté I&C

Week(s ~~6~~ 5 &) ~~7~~ 6 keypoints



Week ~~6~~ 5:

- (related to week ~~5~~ 3 as well) what "lemmatization" is
- what "part-of-speech tagging" is
- two hypothesis to transform PoS tagging into "the second problem" of HMMs
- order of magnitude of performances

Week ~~7~~ 6:

- what an HMM is
- the 3 problems and how it relates to PoS tagging
- Viterbi algorithm
- properties of Baum-Welch algorithm

J.-C. Chappelier & M. Rajman

Input: w_1 . . . . . . w_n

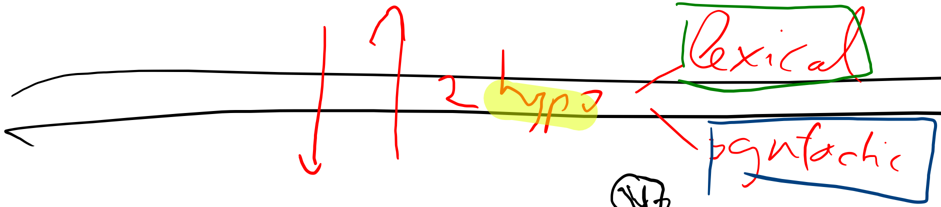
Output: T_1 . . . . . . T_n

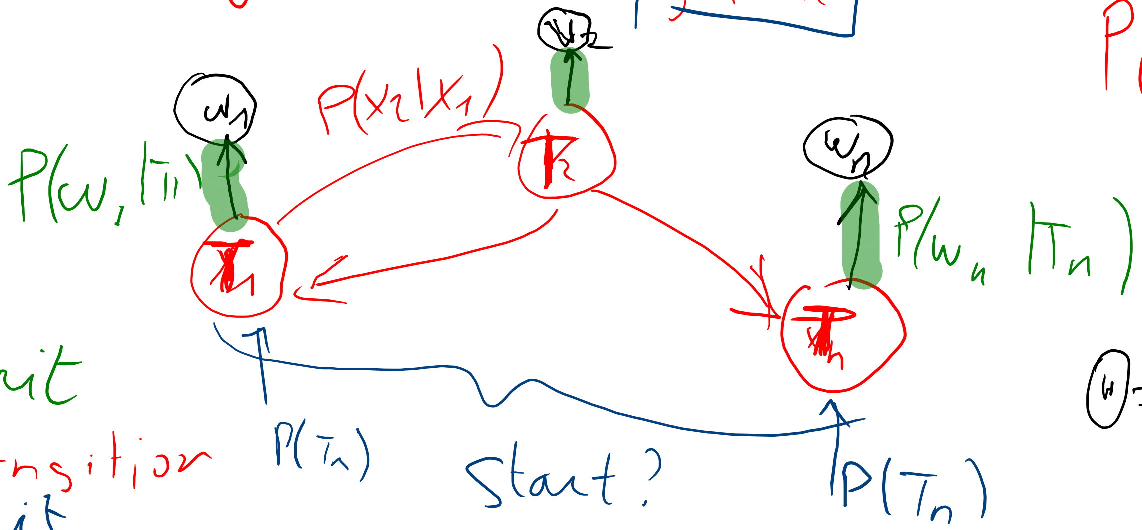
$P(w_i \mid T_1 \cdots T_n w_1 \cdots w_{i-1} w_{i+1} \cdots w_n)$

$= P(w_i \mid T_i)$

$P(T_i \mid T_1 \cdots T_{i-1})$

$= P(T_i \mid T_{i-1})$





$P(x_i \mid x_1 \cdots x_{i-1}) = P(x_i \mid x_{i-1})$

emit

transition

init

$\theta = (P(T_1) \cdots P(T_n)$

$P(T_1|T_2) \cdots$

$P(w_1|T_1) \cdots)$

Input : $w_1 \; w_2 \; \cdots \; w_n$

Output : $T_1 \; T_2 \; \cdots \; T_n$

Lexicon :

$w^{(1)} \; w^{(2)} \; \cdots \; w^{(N)}$

Tag set :

$T^{(1)} \; T^{(2)} \; \cdots \; T^{(M)}$

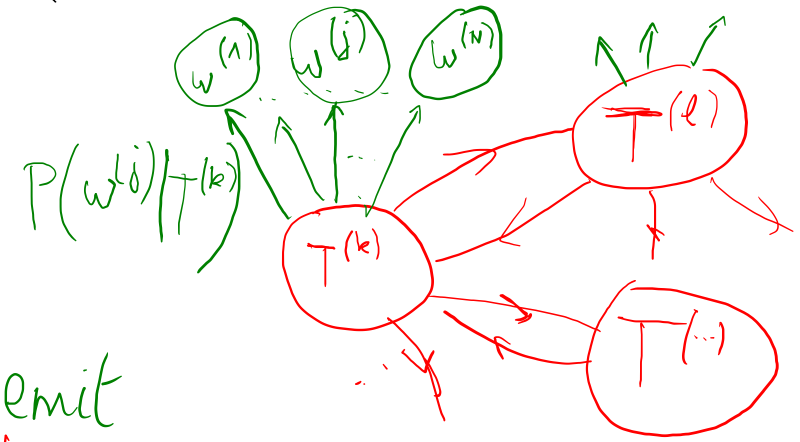
Hypothesis :

lexical :

$$P(w_i \mid T_1 \cdots T_n, w_1 \cdots w_{i-1}, w_{i+1} \cdots w_n) = P(w_i \mid T_i)$$

syntactic :

$$P(T_i \mid T_1 \cdots T_{i-1}) = P(T_i \mid T_{i-1})$$



emit

transition

init

? start ?

$P(T^{(1)}) \cdots\cdots P(T^{(M)})$

$$H = \Big(P(T^{(1)}), P(T^{(2)}), \ldots, P(T^{(M)}),$$
$$P(T^{(1)} \mid T^{(1)}) \; P(T^{(1)} \mid T^{(2)}) \cdots P(T^{(i)} \mid T^{(h)}) \cdots P(T^{(M)} \mid T^{(M)}),$$
$$P(w^{(1)} \mid T^{(1)}) \cdots P(w^{(j)} \mid T^{(k)}) \cdots P(w^{(N)} \mid T^{(M)}) \Big)$$

$$P(X_{pos} = a \mid \text{Universe})$$

$$= P(X_{pos} = a \mid \underbrace{\overset{\text{neighborhood}}{\mathcal{N}}(X_{pos})}_{\text{Markov}})$$

Week 7 practice example (1/3)

① Consider an order-1 HMM PoS tagger using a lexicon with N entries, and a tag set with T tags. Furthermore, assume that the entries of the lexicon are associated, on the average, with t distinct tags.
Provide the total number P of (not necessarily free) parameters to be estimated to exploit the order-1 HMM model, assuming that no guesser has been implemented. **Justify** your answer.

1

2 several

$P =$ $T + T \times T + N \times \{t, T\}$?

free

$(T-1) + T \times (T-1) + T.$ $(N-1)$

② Consider the following lexicon excerpt, where `D`, `N`, `P`, and `V` are the tags associated with the entries
(`D` stands for determiner, `N` for noun, `P` for pronoun, and `V` for verb):

```
cat: N, V                saw: N, V
run: N, V                the: D
running: N, V            you: P
```

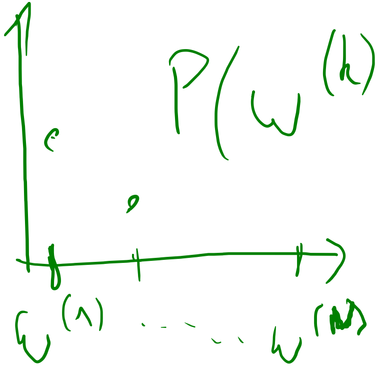
Provide and justify the number M of potential PoS taggings that have to be considered for the following sentence:

`the cat you saw running`

$1 \cdot 2 \cdot 1 \cdot 2 \cdot 2$

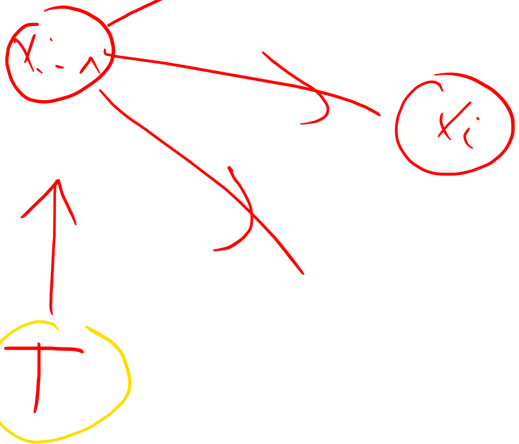
$M =$ 8

J.-C. Chappelier & M. Rajman
EPFL



$P(w^{(k)} \mid X_{i-1})$

$\left. P(X_i \mid X_{i-1}) \right\} T$



$\sum_{x_i} P(X_i \mid X_{i-1}) = 1$

$T \longrightarrow T-1$

$\underbrace{\sum = 1 \quad T}_{T \times (N-1)}$

$w^{(1)}$ ate
$w^{(2)}$ cat
$w^{(3)}$ the
$w^{(4)}$ zulu
} lexicon

the cat ate
$w_1 \quad w_2 \quad w_3$

$w_1 \dots w_n$

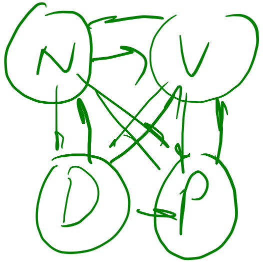
$\underline{\underline{1}}$
pos

$$P(w_i \mid T_i) = P(w_i = w^{(j)} \mid T_i = \boxed{T^{(j)}})$$

Week 7 practice example (2/3)

tag set $T = \{N, V, D, P\}$ $|T| = 4$

$T^{(1)}$ $T^{(2)}$



```
cat: N, V
run: N, V
running: N, V
```

$t = \frac{10}{6}$

```
saw: N, V
the: D
you: P
```

```
the cat you saw running
```

③ What is the condition to be verified by the parameters of the order-1 HMM model (using the provided lexicon excerpt) for the word "`cat`" to be tagged as a noun in the above sentence?
Justify your answer.

J.-C. Chappelier & M. Rajman

Week 7 practice example (2/3)

```
cat: N, V          saw: N, V
run: N, V          the: D
running: N, V      you: P
```

`the cat you saw running`

③ What is the condition to be verified by the parameters of the order-1 HMM model (using the provided lexicon excerpt) for the word "`cat`" to be tagged as a noun in the above sentence?
Justify your answer.

$$P(V|D)\cdot P(cat|V)\cdot P(P|V) < P(N|D)\,P(cat|N)\cdot P(P|N)$$

J.-C. Chappelier & M. Rajman

$$\underset{\text{global max}}{\longrightarrow} \boxed{P(X_1 \dots X_n)}$$

$$= \prod \dots \underbrace{P(X_i \mid X_{i-1})}$$

1-order Markov:
how the global proba is <u>written</u>
(NOT how is it maximized!)

Week 7 practice example (3/3)

T_1	T_2	T_3	T_4	T_n	
D	N	P	V	V	Task
the	cat	you	saw	running	
w_1	w_2			w_n	

④ What is the most probable tagging (using data provided below)?

data
Model

lexicon

$w^{(1)}$

```
cat: N (1e-4), V (2e-6)          saw: N (7e-4), V (8e-5)
run: N (3e-6), V (4e-4)          the: D
running: N (5e-6), V (6e-4)      you: P
```

N

$w^{(w)}$

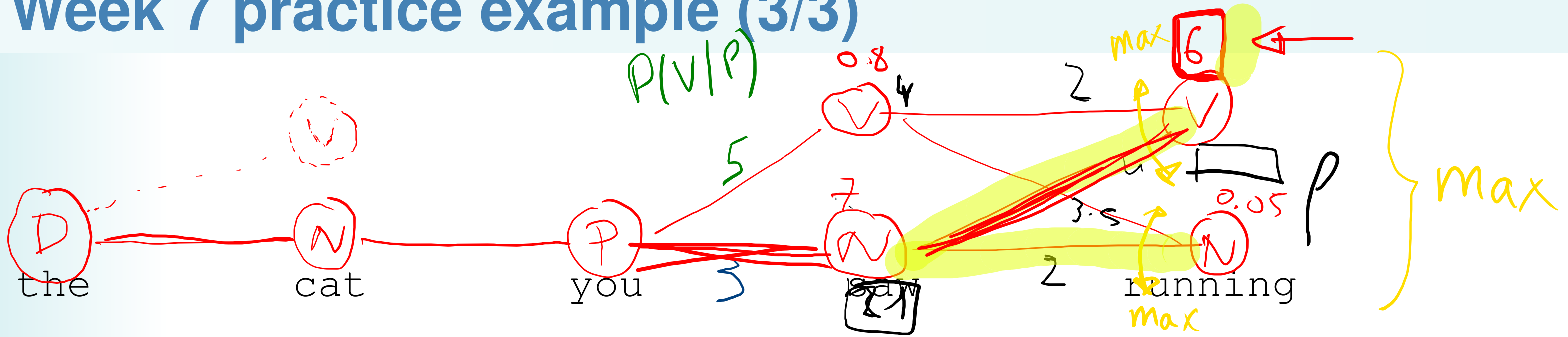
Pi(D) = 0.35	Pi(N) = 0.25	Pi(V) = 0.15	Pi(P) = 0.1

H

P(D|D) = 0	P(N|D) = 0.8	P(V|D) = 0	P(P|D) = 0
P(D|N) = 0.1	P(N|N) = 0.2	P(V|N) = 0.4	P(P|N) = 0.3
P(D|V) = 0.15	P(N|V) = 0.35	P(V|V) = 0.2	P(P|V) = 0.25
P(D|P) = 0.1	P(N|P) = 0.3	P(V|P) = 0.5	P(P|P) = 0

J.-C. Chappelier & M. Rajman

Week 7 practice example (3/3)



④ What is the most probable tagging (using data provided below)?

emit

```
cat: N (1e-4), V (2e-6)          saw: N (7e-4), V (8e-5)
run: N (3e-6), V (4e-4)          the: D
running: N (5e-6), V (6e-4)      you: P
```

Pi(D) = 0.35 Pi(N) = 0.25 Pi(V) = 0.15 Pi(P) = 0.1

D N P N V

P(D|D) = 0	P(N|D) = 0.8	P(V|D) = 0	P(P|D) = 0
P(D|N) = 0.1	P(N|N) = 0.2	P(V|N) = 0.4	P(P|N) = 0.3
P(D|V) = 0.15	P(N|V) = 0.35	P(V|V) = 0.2	P(P|V) = 0.25
P(D|P) = 0.1	P(N|P) = 0.3	P(V|P) = 0.5	P(P|P) = 0

J.-C. Chappelier & M. Rajman