Topic: POS Tagging

$S = \{$今天, 学习, NLP$\}$

$S = w_1\ w_2\ w_3\ w_4\ w_5$

$\quad\ \ z_1\ z_2\ z_3\ z_4\ z_5$

$S' = w_1'\ w_2'\ w_3' \cdots w_N'$

$Z = z_1'\ z_2'\ z_3' \quad z_N' = ?$

$S=$ 今天, 学习, NLP

$S = \underline{<S>}$, 今天 学习 NLP

$$\hat{z} = \arg\max p(z|S)$$

$$= \arg\max \prod_{i=1}^{n} p(w_i|z_i) \cdot p(z_1) \cdot \prod_{t=2}^{n} p(z_t|z_{t-1})$$

$$= \arg\max \log\left(\prod_{i=1}^{n} p(w_i|z_i) \cdot p(z_1) \cdot \prod_{t=2}^{n} p(z_t|z_{t-1})\right)$$

$$= \arg\max \sum_{i=1}^{n} \log p(w_i|z_i) + \log p(z_1) + \sum_{t=2}^{n} \log p(z_t|z_{t-1})$$

②

$$\boxed{\hat{z} = \arg\max_{z} \sum_{i=1}^{n} \log \underbrace{p(w_i|z_i)}_{A} + \log \underbrace{p(z_1)}_{\pi} + \sum_{t=2}^{n} \log \underbrace{p(z_t|z_{t-1})}_{B}}$$

$\theta = \{A, B, \pi\}$

Step 1: Compute A, B, π

Step 2: Viterbi Alg

Noisy Channel Model

$$\underbrace{P(z|S)}_{\text{argmax}} \propto \underbrace{P(S|z)}_{\text{Translation Model}} \cdot \underbrace{p(z)}_{\text{Language Model}}$$

argmax argmax ①

$$= P(w_1 w_2 \cdots w_N | z_1 z_2 \cdots z_N) \cdot p(z_1 z_2 \cdots z_n)$$

$$= \underbrace{\prod_{i=1}^{n} p(w_i|z_i)}_{\text{Emission Prob}} \cdot \underbrace{p(z_1) p(z_2|z_1) p(z_3|z_2) \cdots p(z_n|z_{n-1})}_{\text{LM}}$$

$$p(w_1 w_2 w_3 | z_1 z_2 z_3) \neq p(w_1|z_1) \cdot p(w_2|z_2) \cdot p(w_3|z_3)$$

$\xrightarrow{\text{近似}} p(w_1|z_1) \cdot p(w_2|z_2) \cdot p(w_3|z_3)$

了解: w_i(单词) 负 ⟸ z_i(Tag)

$$\hat{z} = \underset{z}{\arg\max} \sum_{i=1}^{n} \log \underbrace{p(w_i \mid z_i)}_{A} + \log \underbrace{p(z_1)}_{\pi} + \sum_{t=2}^{n} \underbrace{p(z_t \mid z_{t-1})}_{B}$$

w_i：单词　z_i：词性

B（状态转移）

M: # of words

$$A = \begin{array}{c} \\ NNP \\ \vdots \\ VB \\ \vdots \end{array} \begin{bmatrix} \cdots & \cdots & \cdots \\ & & \\ 10^{-6} \; \cdots & 0.015 \; \cdots & 10^{-6} \\ & & \end{bmatrix}$$

apple … study … NNP

词库

N: # of tags

$\boxed{A}$: $\boxed{N \times M}$

$$B = \begin{array}{c|cccc} & NNP & VB & JJ & VBD \; \cdots \\ \hline NNP & 0.02 & 0.03 & 0.07 & 0.1 \\ VB & & & & \\ JJ & & & & \\ VBD & & & & \\ \vdots & & & & \end{array} = 1$$

N

N: # of tags

B: $\boxed{N \times N}$

N: # of tags

π: [| VB | | … | NNP | |]

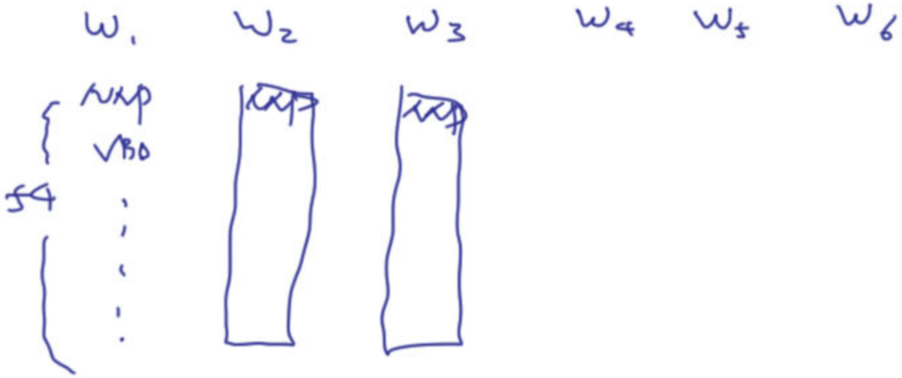
VB → 0.02　NNP → 0.07

VB

NNP

$$\hat{z} = \arg\max \left[\sum_{i=1}^{T} \log p(w_i|z_i) + \log(z_1) + \sum_{t=2}^{T} p(z_t|z_{t-1}) \right]$$

w_1	w_2	w_3	w_4	w_5	w_6
Newsweek	said	it	will	introduce	the
(NNP	VBD	PRP	MD	VB	DT)



$O(54^6$ 组合$) = O(P^n)$

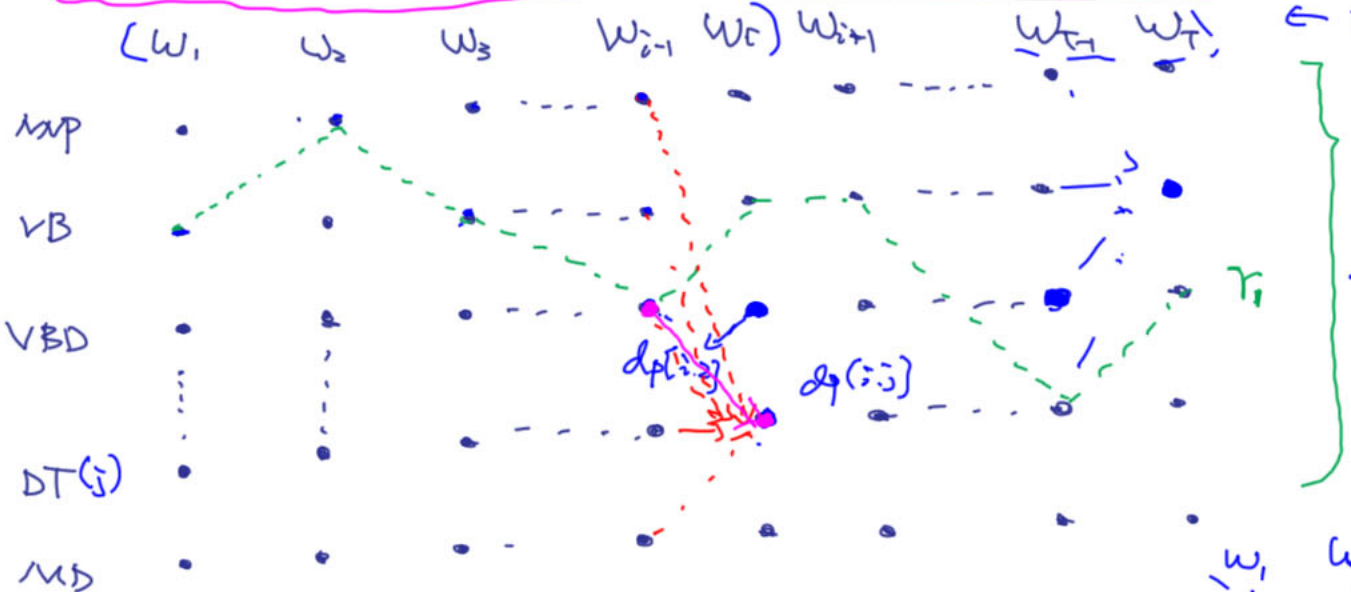
$\Rightarrow$ (NNP, VB, VBD, DT, NNP, NNP)

$$\log p(\text{Newsweek}|\text{NNP}) + \log p(\text{said}|\text{VB}) \cdots + \log p(\text{the}|\text{NNP})$$
$$+ \log p(\text{NNP}) + \log p(\text{VB}|\text{NNP}) + \log(\text{VBD}|\text{VB})$$
$$+ \cdots + \log(\text{NNP}|\text{NNP}) = ✓$$

给定 $w_1, w_2 \cdots w_T$，找出 $z_1 z_2 \cdots z_T$

$\underbrace{w_1, w_2 \cdots w_T}_{\text{单词序列}}$　$\underbrace{z_1 z_2 \cdots z_T}_{\text{词性}}$

$$\boxed{\hat{z} = \arg\max_{z} \sum_{i=1}^{T} \log p(w_i|z_i) + \log p(z_1) + \sum_{t=2}^{T} \log p(z_t|z_{t-1})}$$



54^6 $\text{Score}(r_1) = \log p(VB) + \log p(w_1|VB) + \cdots$ -- w_1

$+ \log p(NNP|VB) + \log p(w_2|NNP)$ -- w_2

$+ \log p(VB|NNP) + \log p(w_3|VB)$ -- w_3

$+ \cdots\cdots$

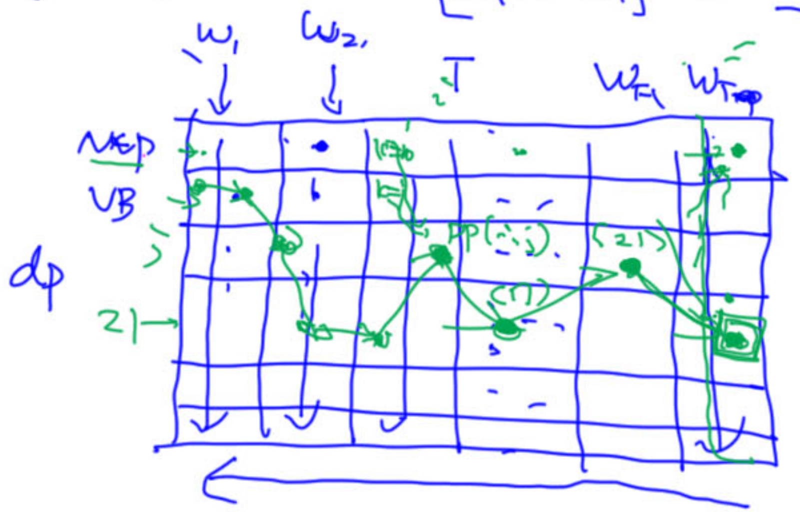
$\boxed{dp[i,j]}$: the best path such that ending at j at time i

$[w_1 \cdots w_i]$，$w_i \leftarrow$ 第j个tag

$$dp[i,j] = \max \begin{cases} dp[i-1,0] + \log p[DT|NNP] + \log p(w_i|DT) \\ dp[i-1,1] + \log p[DT|VB] + \log p(w_i|DT) \\ dp[i-1,2] + \log p[DT|VBD] + \log p(w_i|DT) \\ \cdots\cdots \end{cases}$$

$i = 1 \cdots T$　$j = 0, 1 \cdots 54$

$[w_1 \cdots w_T] \Rightarrow dp[T,0],\ dp[T,1],\ dp[T,2] \cdots dp[T,54]$

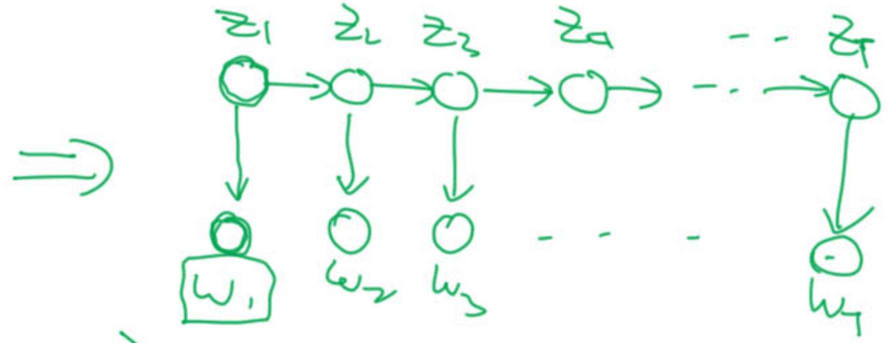


$\boxed{O(T \cdot N \cdot N)}$

$O(N^2 T)$

N: # of tags

T: len(sentence)



$\Rightarrow$ word-h..

目标: Viterbi

Extentions:

① 今天: $D=\{(S_1,z_1),(S_2,z_2)\cdots(S_n,z_n)\}$ 已知 $\boxed{D=\{S_1,S_2,S_3\cdots S_n\}}$

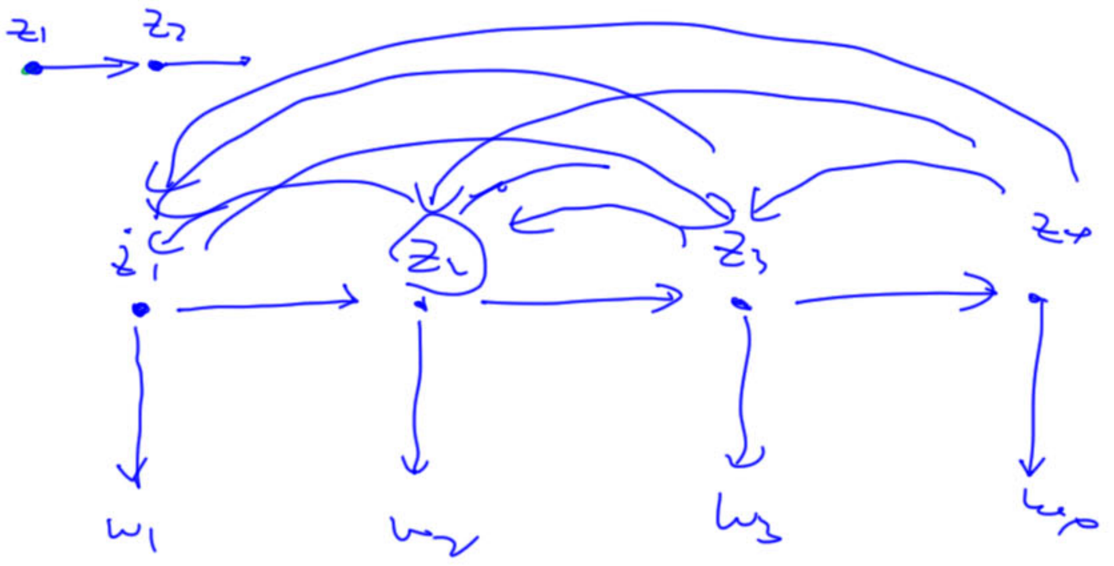
学习时用到 EM 算法

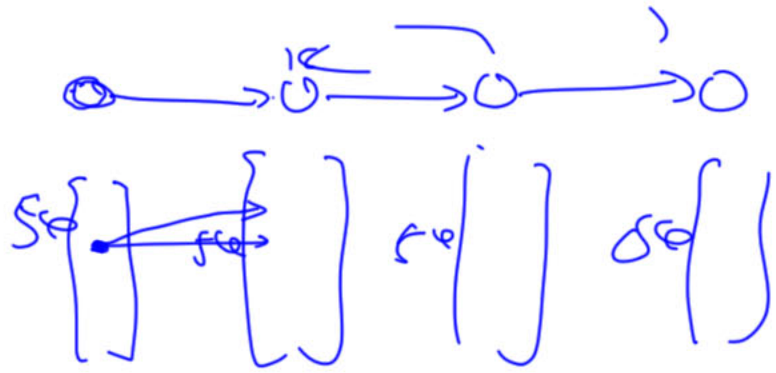
② $P(w_i|z_i) \Rightarrow$ 当前词性与当前单词的关系 $\Rightarrow$ 当前词性与(单词，前面单词，后面单词，前面单词的词性，后面单词词性，当前单词前缀/后缀)

$\Downarrow$

+ CRF 模型

③ BiLSTM + CRF





Latch: = 全局最优解