B58A0050-NLP-Autumn-Homework1 – POS Tagging

58119125 JiangZhuouyang

November 2021

1 Hidden Markov Models and the Viterbi Algorithm

We have a toy language with 2 words - “clean” and “home”. We want to tag the parts of speech in a test corpus in this toy language. There are only 2 parts of speech — NN (noun) and VB (verb) in this language.

We have a corpus of text in which we the following distribution of the 2 words:

	NN	VB
home	8	3
clean	2	7

Assume that we have an HMM model with the following transition probabilities (* is a special start of the sentence symbol).

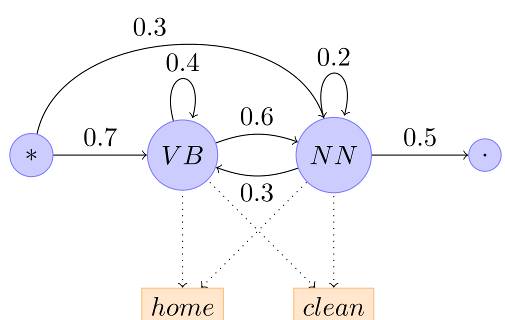

Figure 1: HMM model for POS tagging in our toy language.

(a) Compute the emission probabilities for each word given each POS tag.

(b) Draw the Viterbi trellis for the sequence “clean home.”. Highlight the most likely sequence.

Solution :

(a) $P(home|NN) = 0.8$
$P(clean|NN) = 0.2$
$P(home|VB) = 0.3$
$P(clean|VB) = 0.7$

(b) Let us define the parameter θ of Viterbi trellis:

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

$$\boldsymbol{A} = \begin{array}{c c} & \begin{array}{ccc} NN & VB & . \end{array} \\ \begin{array}{c} NN \\ VB \\ . \end{array} & \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.6 & 0.4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\boldsymbol{B} = \begin{array}{c c} & \begin{array}{ccc} home & clean & . \end{array} \\ \begin{array}{c} NN \\ VB \\ . \end{array} & \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.7 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

$$\boldsymbol{\pi} = \begin{array}{c} \begin{array}{ccc} NN & VB & . \end{array} \\ \begin{pmatrix} 0.3 & 0.7 & 0 \end{pmatrix} \end{array}$$

Then we can use Viterbi Algorithm:

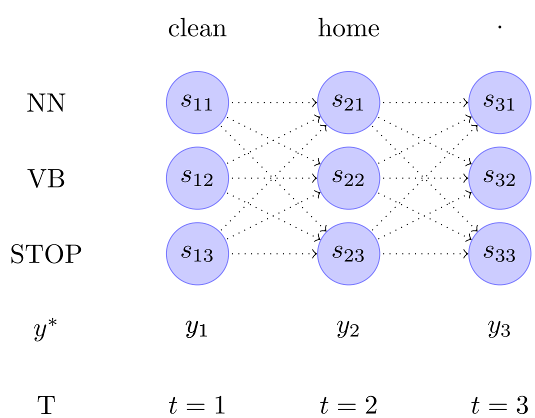

Figure 2: HMM Sequence Mapping.

In such a mapping picture, we can know that, the sequence sorted by 't' is 'clean home .'

Then we calculate each probability item of the sequence:

$$\begin{cases} s_{11} = \pi_1 b_{1,2} = 0.3 * 0.2 = 0.06 \\ s_{12} = \pi_2 b_{2,2} = 0.7 * 0.7 = 0.49 \\ s_{13} = 0 \end{cases}$$

Then we have:

$$\begin{cases} s_{21} = \max_{i=1,2,3}[s_{1i} * a_{i,1}] * b_{1,1} = \max\left[0.012(i=1), 0.294(i=2)\right] * 0.8 = 0.1992(i=2) \\ s_{22} = \max_{i=1,2,3}[s_{1i} * a_{i,2}] * b_{2,1} = \max\left[0.018(i=1), 0.196(i=2)\right] * 0.3 = 0.0588(i=2) \\ s_{23} = 0 \end{cases}$$

Because $s_{12} > s_{11} > s_{13}$ and when $i = 2(VB)$, we have a bigger result from s_{12}
So that, $y_1 = VB$

Than we have:

$$\begin{cases} s_{31} = 0 \\ s_{32} = 0 \\ s_{33} = \max_{i=1,2,3}[s_{2i} * a_{i,3}] * b_{3,3} = \max\left[0.0996(i=1), 0.0294(i=2)\right] * 1 = 0.0996(i=1) \end{cases}$$

Because $s_{21} > s_{22} > s_{23}$ and when $i = 1(NN)$, we have a bigger result from s_{21}
So that, $y_2 = NN$
Because $s_{33} > s_{32} = s_{31}$
So that, $y_3 = STOP$

Finally the result is:

$$y^* = \{VB, NN, STOP\}$$

With a graph, we can represent the sequence:

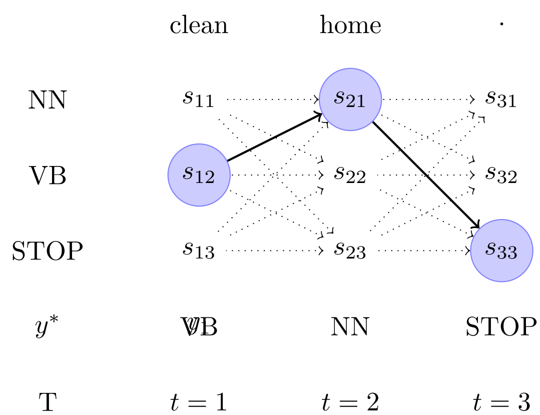

Figure 3: Result Mapping.