NLP Numericals-2

Q1. Using the UPenn Treebank tag set, tag the part of speech to the given statement.

The grand jury commented on a number of other topics.

Determinat - DT
Adjective - JJ
Noun singular - NN
Noun plural - NNS
Verb (Past tense) - VBD
Preposition - IN

Ans

The|DT grand|JJ jury|NN
commented|VBD on|IN a|DT
number|NN of|IN other|JJ
topics|NNS.

Q2. Using HMM POS tagging, tag the following sentence based on the given corpus:

Sentence to be tagged: **Justin will spot Will**

Corpus:

<s> Martin Justin can watch Will</s>

<s>Spot will watch Martin</s>

<s>Will Justin Spot Martin<s>

<s>Martin will pat Spot</s>

POS Tagging
- Rule Based
- Probability / Stochastic

Step2 : Emission Probability Matrix

Bayes Rule $P(word|class) = \frac{P(class, w)}{P(class)}$

Step1 : Assign correct POS tags

<s> Martin Justin can watch Will </s>
N N M V N

<s> Spot will watch Martin </s>
N M V N

<s> Will Justin Spot Martin </s>
M N V N

<s> Martin will pat Spot </s>
N M V N

How many martin considered as Noun / Total Nouns

	N	M	V
Martin	4/9	0	0
Justin	2/9	0	0
Will	1/9	3/4	0
Spot	2/9	0	1/4
can	0	1/4	0
watch	0	0	2/4
pat	0	0	1/4

Step3- State Transition Probability Matrix

	N	M	V	</s>
<s>	3/4	1/4	0	0
N	1/9	1/3	1/9	4/9
M	1/4	0	3/4	0
V	4/4	0	0	0

N gram model,

$$P(W_i|W_{i-1}) = \frac{\text{Count}(W_{i-1}, W_i)}{\text{Count}(W_{i-1})}$$

Probability of Noun after <s> / Total <s>

Step4: Justin will spot Will

N → N, M; N → N, V; N → N, M

1 × 2 × 2 × 2 = 8 ways

According to Viterbi Algorithm

① Justin as Noun

$$P(N|\text{Justin}, \langle s\rangle)$$
$$= P(\text{Justin}|N) \times P(N|\langle s\rangle)$$
$$= \frac{2}{9} \times \frac{3}{4} = \frac{1}{6}$$

② Will as Noun

$$P(N|\text{will}, N)$$
$$= P(\text{will}|N) \times P(N|N)$$
$$= \frac{1}{9} \times \frac{1}{9} = \frac{1}{81}$$
$$= \frac{1}{81} \times \frac{1}{6} = \frac{1}{486}$$

③ Will as Modal

$$P(M|\text{will}, N) = P(\text{will}|M) \times P(M|N)$$
$$= \frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$$
$$= \frac{1}{4} \times \frac{1}{6} = \frac{1}{24}$$

$\because \frac{1}{24} > \frac{1}{486}$

$\therefore$ Will is modal

④ Spot as Noun

$$P(N|\text{spot}, M) = P(\text{spot}|N) \times P(N|M)$$
$$= \frac{2}{9} \times \frac{1}{4} = \frac{1}{18}$$
$$= \frac{1}{18} \times \frac{1}{24} = \frac{1}{432}$$

⑤ Spot as Verb

$$P(V|\text{spot}, M) = P(\text{spot}|V) \times P(V|M)$$
$$= \frac{1}{4} \times \frac{3}{4} = \frac{3}{16}$$
$$= \frac{3}{16} \times \frac{1}{24} = \frac{1}{128}$$

$\frac{1}{128} > \frac{1}{432}$

$\therefore$ spot is verb

⑥ Will as Noun

$$P(N|will, V) = P(will|N) \times P(N|V)$$

$$= \frac{1}{9} \times 1 = \frac{1}{9}$$

$$= \frac{1}{9} \times \frac{1}{128} = \frac{1}{1152}$$

⑦ Will as Modal

$$P(M|will, V) = P(will|M) \times P(M|V)$$

$$= \frac{3}{4} \times \frac{1}{1000} = \frac{3}{4000}$$

$$= \frac{3}{4000} \times \frac{1}{128} = \frac{3}{512000}$$

$\frac{1}{1152} > \frac{3}{512000}$

∴ Will will be Noun

Justin will spot Will

N M V N

Q3. Based on the given state transition and emission probability matrix, assign POS to the statement: **Time flies like an arrow.**

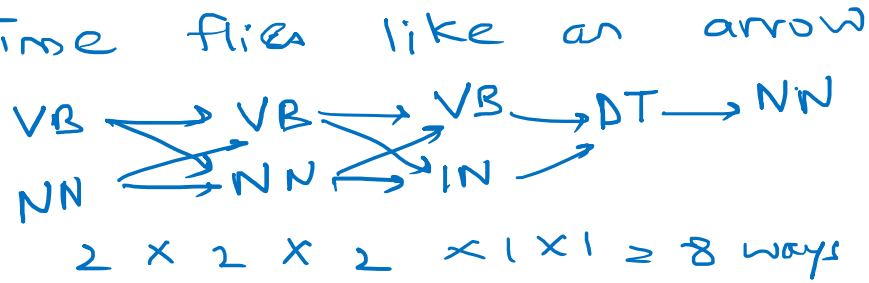
Emission Probability Matrix

	Time	flies	like	an	arrow
VB	0.1	0.2	0.2	0	0
NN	0.1	0.1	0	0	0.1
IN	0	0	0.25	0	0
DT	0	0	0	0.5	0

State Transition Matrix

	VB	NN	IN	DT	</S>
<S>	0.2	0.8	0	0	0
VB	0	0.3	0.2	0.5	0
NN	0.4	0.5	0.1	0	0
IN	0	0.75	0	0.25	0
DT	0	1	0	0	0

Step1 Time flies like an arrow



2 x 2 x 2 x 1 x 1 = 8 ways

Step2 Time as Noun

$P(N|Time, <s>) = P(Time|N) \times P(N|<s>)$

$= 0.1 \times 0.8$

$= 0.08$

Time as verb

$P(V|Time, <s>) = P(Time|V) \times P(V|<s>)$

$= 0.1 \times 0.2$

$= 0.02$

$\therefore 0.08 > 0.02$

$\therefore$ Time will be Noun

flies as Noun

$P(NN|flies, N)$

$= P(flies|N) \times P(N|N)$

$= 0.1 \times 0.5$

$= 0.05 \times 0.08$

$= 0.004$

flies as verb

$P(VB|flies, N)$

$= P(flies|VB) \times P(V|N)$

$= 0.2 \times 0.4$

$= 0.08 \times 0.08$

$= 0.0064$

$0.0064 > 0.004$

$\therefore$ flies is verb

Like as verb

$P(VB|like, V) = P(like|VB) \times P(V|V)$

$= 0.2 \times 0.01$

$= 0.002 \times 0.0064$

$= 0.000128$

Like as Preposition

$P(IN|like, V) = P(like|IN) \times P(IN|V)$

$= 0.25 \times 0.2$

$= 0.50 \times 0.0064$

$= 0.00032$

$0.00032 > 0.000128$

$\therefore$ like is Preposition

Time	flies	like	an	arrow
NN	VB	IN	DT	NN

Q4. Based on the given state transition and emission probability matrix, assign POS to the statement: **That girl smiles.**

Emission Probability Matrix

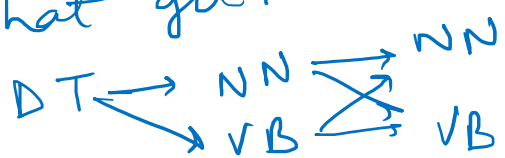
	DT	NN	VB
That	0.4	0	0
girl	0	0.015	0.0031
smiles	0	0.0004	0.2

State Transition Matrix

	DT	NN	VB
<S>	0.5	0.4	0.1
DT	0.01	0.99	0
NN	0.3	0.3	0.4
VB	0.4	0.4	0.2

Step1

That girl smiles



1 X 2 X 2 = 4 ways

① That as DT

$P(DT|That, <s>)$

$= P(that|DT) \times P(DT|<s>)$

$= 0.4 \times 0.5$

$= 0.2$

② Girl as Noun

$P(Noun|Girl, DT)$

$= P(Girl|N) \times P(N|DT)$

$= 0.015 \times 0.99$

$= 0.0148 \times 0.2$

$= 0.0029$

④ Girl as VB

$P(VB|girl, DT)$

$= P(girl|VB) \times P(VB|DT)$

$= 0.0031 \times 0.001$

$= 0.00031 \times 0.2$

$= 0.00062$

$0.0029 > 0.00062$

∴ Girl is a Noun

⑤ Smiles as NN

$P(N|smiles, N)$

$= P(smiles|N) \times P(N|N)$

$= 0.0004 \times 0.3$

$= 0.00012$

$= 0.00012 \times 0.003$

$= 0.00000036$

⑥ Smiles as VB

$P(smiles|V) \times P(V|N)$

$= 0.2 \times 0.4$

$= 0.08 \times 0.003$

$= 0.00024$

∴ Smile is verb.

That girl smiles

DT NN VB

Q5. Using HMM POS tagging, tag the following sentence based on the given corpus:

Sentence to be tagged: **The Park is a book.**

Corpus:

<s> Book a car</s>

<s>Park the car</s>

<s>The book is in the car</s>

<s>The car is in a park</s>

Step1 : Apply correct tag

<s> Book a car </s>
V D N

<s> Park the car </s>
V D N

<s> The book is in the car </s>
D N D IN D N

<s> The car is in a park </s>
D N D IN D N

Step1 Emission Probability Matrix

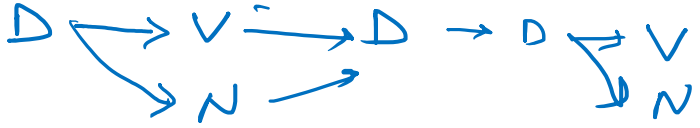
	V	D	N	IN
Book	1/2	0	1/6	0
a	0	2/8	0	0
car	0	0	4/6	0
Park	1/2	0	1/6	0
the	0	4/8	0	0
is	0	2/8	0	0
in	0	0	0	2/2

Step3 State Transition Matrix

	V	D	N	IN	<s>
<s>	2/4	2/4	0	0	0
V	0	2/2	0	0	0
D	0	0	4/6	2/6	0
N	0	2/6	0	0	2/6
IN	0	2/2	0	0	0

Step4 Test Data

The Park is a book



1 × 2 × 1 × 1 × 2 = 4 ways

Step5

① The as Determinant

$P(D|The, <s>)$

$= P(The|D) \times P(D|<s>)$

$= \frac{4}{8} \times \frac{1}{4}$

$= \frac{1}{4}$

② Park as Verb

$P(V|Park, D)$

$= P(Park|V) \times P(V|D)$

$= \frac{1}{2} \times \frac{1}{1000} \times \frac{1}{4}$

$= \frac{1}{8000}$

③ Park as Noun

$P(N|Park, D)$

$= P(Park|N) \times P(N|D)$

$= \frac{1}{6} \times \frac{4}{6}$

$= \frac{1}{9} \times \frac{1}{4} = \frac{1}{36}$

$\because \frac{1}{36} > \frac{1}{8000}$

∴ Park will be Noun

④ is as a determinant

$P(D|is, N) = P(is|D) \times P(D|N)$

$= \frac{2}{8} \times \frac{2}{6}$

$= \frac{1}{12} \times \frac{1}{36}$

$= \frac{1}{432}$

(5) a as DT

$P(D|a, D)$

$= P(a|D) \times P(D|D)$

$= \frac{2}{8} \times \frac{1}{1000}$

$= \frac{2}{8000} = \frac{1}{4000} \times \frac{1}{432} = \frac{1}{1728000}$

(6) Book as Noun

$P(N|Book, D)$

$= P(Book|N) \times P(N|D)$

$= \frac{1}{6} \times \frac{4}{6} = \frac{1}{9} \times \frac{1}{172800}$

$= \frac{1}{1555200}$

(7) Book as Verb

$P(V|Book, D)$

$= P(Book|V) \times P(V|D)$

$= \frac{1}{2} \times \frac{1}{1000} \times \frac{1}{1728000}$

$= \frac{1}{3456 \times 10^6}$

∴ Book is Noun

∴ The Park is a Book

DT NN ~~DT~~ DT NN