ID :13662
Syed Muhammad salman khan
Submitted to : Aasma khan
Summer Assigment

Sessional (Summer- 2020)

Natural Language Processing

Note: Attempt all Questions.

Question No. 1: **(15)**

a) Explain Part of Speech Tagging (POS) and explain POS tag ambiguity with two examples.

Part of speech tagging is also known as part of speech tags ,lexical categories, word classes morphological classes , lexical tags etc.

The process of assigning a part-of-speech to each word in a sentence.

Example :play well with other.

Play	V
Well	Ad
With	Prep
Other	N

Example :

Heat	Verb(noun)
Water	Noun(verb)
In	Prep(noun ,adj)
A	Det (noun)
Large	Adj(noun)
Vessel	noun

(03)

b) State difference between open vs. closed classes.

Open classes : unlimited numbers of words
Open classes allow new members through borrowing (for example, the noun *cafe*) and derivation (for example, the adjective *bounteous* from the noun *bounty*)

Examples : Noun , verb ,Adverb , Adjective…

Closed classes: Closed classes of words do not allow new members and usually involve grammatical rather than lexical words

Examples : Auxiliary , Articles, Determine ,conjunction ,pronoun ,interjections…
(02)

c) Apply Viterbi Algorithm on the below given bigram and lexical probabilities; (10)

Initial Probabilities	
Noun	1\3
Verb	0
Other	1\3

Bigram Probabilities			
	Noun	Verb	Other
Noun	1\4	1\4	0
Verb	1\4	0	1\4
Other	1\3	0	1\3

Lexical Probabilities					
	O1=time	O2=flies	O3=like	O4=an	O5=arrow
Noun	1\5	1\5	0	0	1\5
Verb	1\5	2\5	1\5	0	0
Other	0	0	1\5	2\5	0

Lexical probability

	Time	**Flies**	**Like**	**An**	**Arrow**
Noun	**1/5**	**1/5**	**0**	**0**	**1/5**
Verb	**1/5**	**2/5**	**1/5**	**0**	**0**
Other	**0**	**0**	**1/5**	**2/5**	**0**

	P1(h1)	P2(h2)	P3(h3)	P4(h4)	P5(h5)
Noun	**1/2*1/5**	**1/2.1/5.1/3.1/5=1/150**	**0**	**0**	**1/3750.1/3.1/5=1/562500**
Verb	**0**	**1/2.1/5.1/3.2/5=1/75**	**1/75.1/3.1/5=1/1125**	**0**	**0**
Other	**0**	**0**	**1/75.1/5.1/2=1/750**	**1/750.1/2.2/5=1/3750**	**0**

Times = Noun
Flies = Verb
Like = Other

Question No. 2: **(05)**

Apply Bayesian theorem over the below given string:
^John got many NLP books. ^He found them all very interesting.
Where for lexical probabilities assume John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.

^John	Got	Many	Nlp	Books .	^he	Found	Them	Very	Interesting.
NN	VB	A	N	N	N	V	N	R	A

	^	N	V	A	R	.
^	0	2	0	0	0	0
N	0	1	2	0	1	1
V	0	1	0	1	0	0
A	0	1	0	0	0	1
R	0	0	0	1	0	0
.	1	0	0	0	0	0

STEP NO 3:

	^	N	V	A	R	.
^	0	2/2=1	0	0	0	0
N	1	1/5	2/5	0	1/5	1/5
V	0	1/2	0	1/2	0	0
A	0	1/2	0	0	0	1/2
R	0	0	0	1/1=1	0	0
.	1/2	0	0	0	0	0

	0.5	0.3	0.2	0.1	0.6
	JOHN	GOT	MANY	NLP	BOOKS
^	0.5	0.3	0.2	0.1	0.6
N	0.5	0.3	0.2	0.1	0.6
V	0.5	0.3	0.2	0.1	0.6
A	0.5	0.3	0.2	0.1	0.6
R	0.5	0.3	0.2	0.1	0.6
.	0.5	0.3	0.2	0.1	0.6

P(JOHN/^) JOHN*^=0.5*2/2=0.5
P(GOT/^) GOT*^ = 0.3*2/2= 0.3
P(MANY/^)MANY*^ = 0.2 *2/2=0.2
P(NLP/^)NLP*^=0.1*2/2=0.1
P(BOOKS/^)BOOKS*^ = 0.6 *2/2=0.6
P(john/N)JOHN*N=0.5*5/5=0.5
and so on...

Good Luck ☺