Department of Computer Science (BS-SE)

IQRA NATIONAL UNIVERSITY PESHAWAR

Sessional (Fall- 2020)

Natural Language Processing

Name	HOORIA KHAN ORAKZAI
ID	14263

Semester: 5th

Deadline: 10th June 2020 **Total Marks: 20**

Instructor: Aasma Khan **Date: 9th May, 2020**

Note: Attempt all Questions.

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

a) Explain Part of Speech Tagging (POS) and explain POS tag ambiguity with two examples.
(03)

- **POS Tagging**: It is a process of converting a sentence to forms – list of words, list of tuples (where each tuple is having a form (word, tag)). The tag in case of is a part-of-speech tag, and signifies whether the word is a noun, adjective, verb, and so on.
- **Types of POS Tagger**: POS-tagging algorithms fall into two distinctive groups: Rule-Based POS Taggers and Stochastic POS Taggers.
- **Default tagging**: is a basic step for the part-of-speech tagging. It is performed using the Default Tagger class. The Default Tagger class takes 'tag' as a single argument. NN is the tag for a singular noun. Default Tagger is most useful when it gets to work with most common part-of-speech tag. That's why a noun tag is recommended.
- **POS tag ambiguity:**
  Common parts of speech in English are noun, verb, adjective, adverb, etc. The POS tagging problem is to determine the POS tag for a particular instance of a word. The main problem with POS tagging is ambiguity. In English, many common words have

multiple meanings and therefore multiple POS. The job of a POS tagger is to resolve this ambiguity accurately based on the context of use.

For Example:

People jump high
People noun/verb
Jump noun/verb
High noun/adjective
List of all possible tag for each word

b) State difference between open vs. closed classes. (02)

Open Classes:

An open class is one that commonly accepts the addition of new words open classes (like nouns, verbs and adjectives) acquire new members constantly. Open classes normally contain large numbers of words

Close Classes:

An closed class is one to which new items are very rarely added, a closed. Open classes (like nouns, verbs and adjectives) acquire new members constantly. Closed classes normally contain small numbers of words

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

hi	P1(h1)	P2(h2)	P3(h3)	P4(h4)	P5(h5)
noun	1/3x1/5=1/15	1/5x1/4x1/15=1/300	0x1/150=0	0	1/5x1/4x1/16875
verb	0	2/5x1/4x1/15=1/150	1/5x1/4x1/150=1/3000	1/3x2/5x1/2250=1/16875	0
other	0	0	1/5x1/3x1/150=1/2250	0	0
	time=noun	flies=verb	like=other	an=other	arrow=noun

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.

Good Luck ☺

ANSWER NO 2:

POS Tags:

^N V A N N. ^ N V N A R A.

Recording Numbers:

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

Bigram Probability:

Bigram Probability= P(x|y)=P(a.b)/P(a)

	^	N	V	A	R	.
^	0	2	0	0	0	0
N	0	1/5	2/5	1/5	0	1/5
V	0	1/2	0	1/2	0	0
A	0	1/3	0	0	1/3	1/3
R	0	0	0	1	0	0
.	1	0	0	0	0	0

Now putting all the values in Bayes theorem i.e.
P(T)=P(W/T)=TTP(Ti-/Ti-1)x P(Wi/Ti)
Where lexical probability is given i.e.
John=0.5, got=0.3, many=0.2, NLP=0.1 and books=0.