Natural Language Processing
Assignment 11
Type of Question: MCQ

Number of Questions: 8 [Question 4,5 carries two marks] Total Marks: 6*1+2*2=10

Question 1: Your teacher recommended you to read the book 'Deep Learning with Python'. After reading the book, you want to summarize it. What kind of summarization method would you use for this purpose?

1. Abstractive single document summarization
2. Abstractive multi document summarization
3. Extractive single document summarization
4. Extractive multi document summarization

a. 1, 2
b. 3, 4
c. 1, 3
d. 2, 4

Answer: c
Solution:

Question2: Which of the following is/are True?

1. PageRank based algorithm is used to compute the sentence centrality vector
2. Query-focused summarization can be thought of as a simple question answering system
3. The underlying hypothesis of LexRank algorithm is sentences that convey the theme

of the document are more similar to each other

4. All of the above

Answer: 1,3

Solution: Refer Lecture 51

Question 3 : Which of the following is/are True?

1. ROUGE metric is as good as human evaluation
2. LexRank can be applied for multi-document summarization
3. In optimization based approach for summarization, the inference problem is to select a subset S of textual units found such that summary score is maximized
4. Maximum Marginal Relevance strives to reduce redundancy while maintaining query relevance.

Answer: 2,3,4

Question 4: It is estimated that 20% of GPT-4 generated texts are fake. Google built some AI system to filter these fake contents. An AI system claims that it can detect 99% of fake contents, and the probability for a false positive (a real content detected as fake) is 3%. Now if a content is detected as fake, then what is the probability that it is in fact a real content?

a. 0.084
b. 0.118
c. 0.108
d. None of the above

Answer: c

Solution:

Let, A = Event that a content is detected as fake
B = Event that a generated text is fake
$P(B) = 0.2$

$P(B') = 0.8$
$P(A|B) = 0.98$
$P(A|B') = 0.03$

$$P(B'|A) = P(A|B')P(B')/P(A)$$

$$= P(A|B')P(B')/(P(A|B)P(B) + P(A|B')P(B'))$$

$$= (0.03 \times 0.8)/(0.99 \times 0.2 + 0.03 \times 0.8)$$

$$\approx 0.108$$

For question 5-8 follow the below Table . One classifier predicts the following. The correct prediction is shown by the tick mark under Match column.

No	Actual	Predicted	Match
1	Airplane	Airplane	✓
2	Car	Boat	
3	Car	Car	✓
4	Car	Car	✓
5	Car	Boat	
6	Airplane	Boat	
7	Boat	Boat	✓
8	Car	Airplane	
9	Airplane	Airplane	✓
10	Car	Car	✓

Question 5: What is the macro-averaged f1 score?

a. 0.54
b. 0.56
c. 0.58
d. 0.64

Answer: c

Question 6: What is the micro averaged precision?

a. 0.58
b. 0.64
c. 0.50
d. 0.60

Answer: d

Question 7: What is the f1 score of boat class?

a. 0.40
b. 0.30
c. 0.58
d. 0.67

Answer: a

Question 8: What is the accuracy of the classifier?

a. 0.40
b. 0.50
c. 0.60
d. 0.90

Answer: c

Solution:

			Predicted	
	Label	**Airplane**	**Boat**	**Car**
	Airplane	**2**	1	0
Actual	**Boat**	0	1	0
	Car	1	2	**3**

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)
Airplane	2	1	1
Boat	1	3	0
Car	3	0	3

Airplane :

Precision: 0.67, Recall: 0.67

Boat:

Precision: 0.25, Recall: 1.00

Car:

Precision: 1.00, Recall: 0.50

macro-f1: 0.58

micro-f1: 0.60

Accuracy: 6/10 = 0.60