NPTEL
NATIONAL PROGRAM ON TECHNOLOGY ENHANCED LEARNING LOCAL CHAPTER

LIVE SESSION 11
NPTEL NLP
(NOC24_CS39)

Shubhi Bansal

PMRF Scholar

IIT Indore

Q1

NPTEL
NATIONAL PROGRAM ON TECHNOLOGY ENHANCED LEARNING LOCAL CHAPTER

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 methods would you use for this purpose?

A

- ❑ Abstractive single document summarization
- ❑ Abstractive multi document summarization
- ❑ Extractive single document summarization
- ❑ Extractive multi document summarization

(a)	1, 2
(b)	3, 4
(c)	1, 3
(d)	2, 4

Q1

NPTEL
NATIONAL PROGRAM ON TECHNOLOGY ENHANCED LEARNING LOCAL CHAPTER

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

Q1

- Ans: c

Q1

- Why Abstractive Summarization is the Best Fit:
  - Comprehensiveness: Technical books like "Deep Learning with Python" cover complex concepts. An abstractive summary allows you to rephrase and synthesize information for better understanding, rather than just extracting verbatim sentences.

    generate (summarize)
  - Conciseness: Abstractive summarization forces you to condense information. This is important for a book where you want to capture the essence without a lengthy summary.
  - Focus on Key Concepts: You can focus on summarizing the most important techniques, algorithms, and insights from the book, even if they aren't stated directly in a single sentence.
- Why Extractive Might Be Considered Less Ideal as a Standalone Method:
  - Technical Jargon: Extracting sentences directly might result in a summary filled with complex terms that are less understandable without the surrounding context a book provides.
  - Missed Connections: Extractive summarization might overlook important connections and relationships between concepts that are explained across several paragraphs or chapters.
- Important Note:
  - While abstractive summarization is generally the more powerful method for this task, it's worth noting that sometimes a combination of abstractive and extractive techniques can be the most effective approach, even if the answer key specifies only one.

generation retrieval

Q2

Q A

What kind of summarization approach is lexrank?

a. Extractive multi document generic (general)

b. Extractive multi document query specific

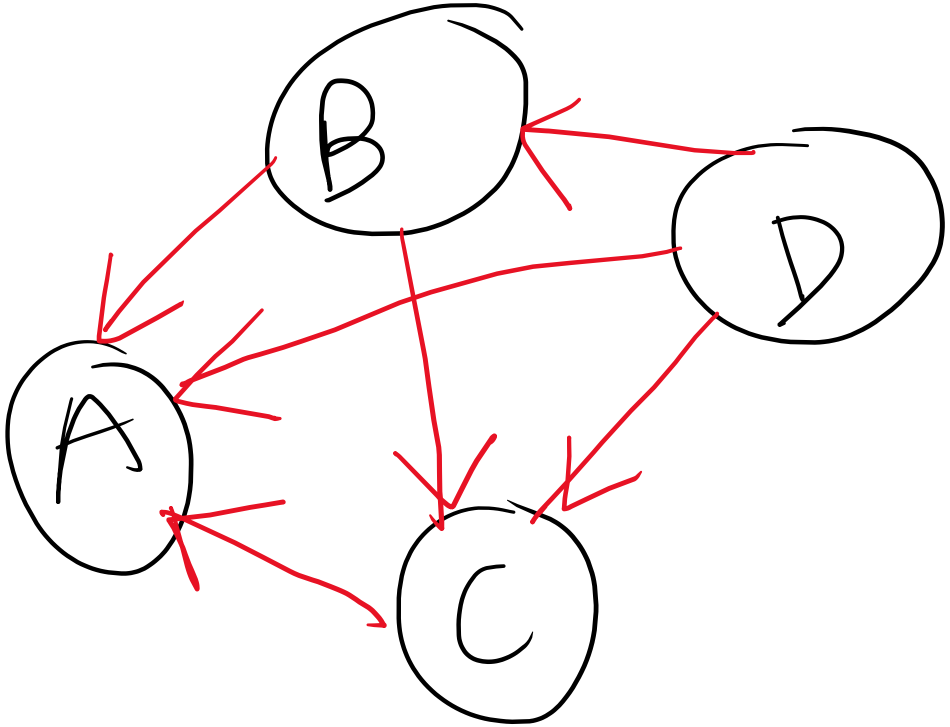
c. Abstactive multi document query specific QA

d. Abstractive multi document generic

LexRank → single
→ multi document
summarization

(derived from PageRank) developed by Sergey Brin Larry Page

- Powers google's search ability



outbound links

Q2

- Ans: a
- The correct answer is a. Extractive multi-document generic. Here's why:
- Extractive: LexRank identifies and selects the most important sentences from the original text(s) to form the summary. This means it doesn't generate new text.
- Multi-document: LexRank is designed to handle multiple documents, allowing you to summarize information from a collection of related texts.
- Generic: LexRank is a generic summarization approach. It's not specifically designed to answer a particular query, but rather to provide a general overview of the important information within the documents.

Q3

Identify whether the following statements are True or False.

1. Maximum Marginal Relevance strives to reduce redundancy while maintaining query relevance.

2. Query-focused summarization can be thought of as a complex question-answering system.

a. True, False

b. True, True

c. False, True

d. False, False

Q3

- Answer: b
- True, True
- Maximum Marginal Relevance (MMR): True. MMR is a classic summarization and re-ranking algorithm focused on the balance between:
  - Relevance: Ensuring the summary answers the user's query or represents the document’s main topics.
  - Novelty/Diversity: Avoiding redundancy by ensuring each selected sentence adds new information.
- Query-focused summarization: True. Query-focused summarization directly addresses a user's question. The goal is to extract the most relevant information and present it as a concise, informative summary, much like how a complex QA system would function.

Questions 4-8

V. Imp

For question 4-8, use the data given in Table 1.

- Suppose you have trained an image classifier with five classes - cat, dog, lion, tiger, and deer.
- Consider the confusion matrix shown in Table 1.

2 classes Binary

5 classes (Multi-class classification)

Actual — Gold Labels

Predicted Labels	cat	dog	lion	tiger	deer
cat	120	17	9	7	40
dog	15	150	25	10	7
lion	10	45	150	23	5
tiger	15	15	20	120	30
deer	40	30	20	10	155

47
73

cat

gold →
Predicted ↓

	TP	FP
	FN	TN

predicted →
gold ↓

	TP	FN
	FP	TN

Q4

- What is the macro averaged precision?

a. 0.6696

b. 0.6078

c. 0.6433

d. None of the above

$$= \frac{0.6404 + 0.7246 + 0.6438 + 0.6 + 0.6078}{5}$$

$$= \boxed{0.64333}$$

"Cat"

TP	FP
FN	TN

Gold / Ground-truth label / Actual : cat

Predicted : Non-cat

TP

Ground/Actual = Predicted ⇒ 130 TP

gold "cat" 'cat'

"cat" ≠ Predicted ⇒ 80

FN

ground truth cat

Predicted not cat

-ve not cat

model wrongly predicts -ve case

TP:130

FN:80

FP:73

TN:815

"cat"

FP

ground truth not cat ≠ Predicted ~~not~~ cat ⇒ 73

(model wrongly predicts +ve case)

FP

model wrongly predicts (+ve) case

true case : 'cat'

~~wrong~~ pred = cat

wrongly ⇒ Pred ≠ GL

FN

model wrongly predicts (-ve) case

GL ≠ Pred

cat = Not cat

Q4

- Ans: c

TN: model 'correctly' predicts '-ve' class

GT = Pred

Not cat = Not cat

Gold

Pred

TP	FP
FN	TN

dog

Pred

A=D, P=D Gold

TP 150	FP 57
FN 107	TN 784

lion

TP 150	FP 83
FN 74	TN 791

Gold Labels

Predicted Labels	cat	dog	lion	tiger	deer
cat	130	17	9	7	40
dog	15	$\boxed{150}$	25	10	7
lion	10	45	$\boxed{150}$	23	5
tiger	15	15	20	$\boxed{120}$	30
deer	40	30	20	10	$\boxed{155}$

FP Σ row $-150 = 57$

Σ sum -150 = FN

Table 1

tiger

120	80
50	848

deer

155	100
82	761

$$Pre = \frac{TP}{TP+FP} =$$

Pred

	GT
TP	FP
FN	TN

cat

TP	FP
130	73
FN 80	TN 815

$$P_c = \frac{130}{130+73}$$

dog

TP	FP
150	57
FN 107	TN 784

$$P_D = \frac{150}{150+57}$$

lion

150	83
74	791

$$P_L = \frac{150}{150+83}$$

tiger

120	80
50	848

$$P_T = \frac{120}{120+80}$$

deer

155	100
82	761

$$P_{de} = \frac{155}{155+100}$$

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN}$$

$$F1 = \frac{2 * P * R}{P+R}$$

Macro-averaged precision =

$$\frac{P_{c1} + P_{c2} + P_{c3} + P_{c4} - - - + P_{cn}}{n}$$

Q5

What is the macro-averaged recall?

a. 0.6464

b. 0.6540

c. 0.6190

d. None of the above

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\frac{0.6190 + 0.5837 + 0.6696 + 0.7059 + 0.6540}{5}$$

$$\frac{R_{c_1} + R_{c_2} + R_{c_3} + R_{c_4} + R_{c_5}}{5} = 0.$$

Q5

- Ans: a

Q6

What is the accuracy of your classifier?

a. 0.6421

b. 0.6536

c. 0.6319

d. None of the above

Pooled overall confusion matrix

TP	FP
FN	TN

$$ACC = \frac{705}{1098} \simeq 0.6421$$

TP: 130 + 150 + 150 + 120 + 155 =

Q6

NPTEL
NATIONAL PROGRAM ON TECHNOLOGY ENHANCED LEARNING LOCAL CHAPTER

- Ans: a

Q7

73+57 +83 +80 +100

TP : 705

overall

What is micro-averaged precision?

micro

a. 0.6915

b. 0.6421

c. 0.6245

d. None of the above

micro ~~use~~ the pooled confusion matrix

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{705}{705+393} \approx 0.6421$$

Q7

- Ans: b

Q8

What is micro-averaged recall?

a. 0.6190

b. 0.6535

c. 0.6421

d. None of the above

Q8

- Ans: c

Q9-12

- For questions 9-12, follow the below table. One classifier predicts the following. The tick mark shows the correct prediction under Match column. Actual →

GT

pred ↓	airplane	car	boat
air	2	1	0
car	0	3	0
boat	1	2	1

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	✓

~~Predi~~

Actual	Pred
A	C
B	C
A	B
C	B

Q9

- What is the macro-averaged F1-score?

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

Q10

- What is the micro-averaged precision?

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

Q11

- What is the f1-score of boat class?

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

Q12

- What is the accuracy of the classifier?

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

Q13

- It is estimated that 20% of GPT-4 generated texts are fake. Google built some AI systems to filter these fake content. An AI system claims that it can detect 98% of fake content, and the probability of a false positive (real document detected as fake) is 3%. Now, if a content is detected as fake, then what is the probability that it is in fact real content?

a. 0.084

b. 0.109

c. 0.119

d. None of the above

Q13

- Ans: b

Q14

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

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

Q14

- p_fake_content = 0.2 (Probability of content being fake)
- p_ai_correct = 0.99 (Probability of AI correctly detecting fake content)
- p_ai_false_positive = 0.03 (Probability of AI incorrectly detecting real content as fake)
- Bayes' Theorem:
- We want to find P(real | detected fake). Bayes' theorem gives us:
- P(real | detected fake) = (P(detected fake | real) * P(real)) / P(detected fake)
- Calculate components:
- P(detected fake | real) = p_ai_false_positive = 0.03
- P(real) = 1 - p_fake_content = 0.8
- P(detected fake): This is the total probability of a document being flagged as fake, considering both true and false positives: P(detected fake) = (p_fake_content * p_ai_correct) + ((1 - p_fake_content) * p_ai_false_positive) = (0.2 * 0.99) + (0.8 * 0.03) = 0.222
- Apply Bayes' theorem:
- P(real | detected fake) = (0.03 * 0.8) / 0.222 = 0.108108... (approximately 0.108)
- Therefore, if a piece of content is detected as fake, there's approximately a 10.8% chance that it's actually real.

Q15

Consider the system-generated summary (S) and the reference summary as follows:

S: ChatGPT is powered by deep learning, a technique that involves training a neural network with extensive data.

R: ChatGPT is a deep learning model that uses a neural network to understand language patterns.

- What is the ROUGE-1 recall for the given summary with respect to the reference?

a. 0.500
b. 0.571
c. 0.470
d. None of the above

Q15

- Ans: b

Q15

- Identify common unigrams:
- System (S): "ChatGPT", "is", "powered", "by", "deep", "learning", "a", "technique", "that", "involves", "training", "a", "neural", "network", "with", "extensive", "data"
- Reference (R): "ChatGPT", "is", "deep", "learning", "model", "that", "uses", "a", "neural", "network", "to", "understand", "language", "patterns"
- Common Words: "ChatGPT", "is", "deep", "learning", "a", "neural", "network“,”that”
- Calculate Recall
- Recall = (Number of common unigrams) / (Number of unigrams in the reference summary)
- Recall = 8 / 14 = 0.571
- Calculate ROUGE-1 Recall Score
- Since precision and recall are typically equally weighted in the final F1 score, the ROUGE-1 recall score is also
- Important Note: While ROUGE-1 recall provides a useful metric, it's important to remember that it doesn't give a complete picture of summary quality. Other ROUGE metrics (like ROUGE-2, which focuses on bigrams) and human evaluation are often necessary to get a more comprehensive assessment.

TextRank vs LexRank

- https://blog.peiyingchi.com/2019/10/29/TextRank-LexRank-DivRank/

Similarity measures

- TextRank: the number of words two sentences have in common normalized by the sentences' lengths

$$Similarity(S_i, S_j) = \frac{|w_k | w_k \in S_i \& w_k \in S_j|}{log(|S_i|) + log(|S_j|)}$$

- LexRank: cosine similarity of TF-IDF vectors:

$$\text{idf-modified-cosine}(x, y) = \frac{\sum_{w \in x,y} \text{tf}_{w,x}\text{tf}_{w,y}(\text{idf}_w)^2}{\sqrt{\sum_{x_i \in x}(\text{tf}_{x_i,x}\text{idf}_{x_i})^2} \times \sqrt{\sum_{y_i \in y}(\text{tf}_{y_i,y}\text{idf}_{y_i})^2}}$$