LIVE SESSION 12 NPTEL NLP (NOC24_CS39)

Shubhi Bansal

PMRF Scholar

IIT Indore

Q1

□ Which of the following are indicators of Irrealis moods?

a. Words enclosed in quotes.

b. Conditional markers such as "If".

c. Questions.

d. All of the above (Ans.)

Q1

- Answer: 4

Q1

- Here's the breakdown of why the answer is All of the above:

- *(b)* Conditional markers such as "if": Conditional sentences ("If I were you...", "If I had more time...") inherently express circumstances that are hypothetical or contrary to fact, marking the irrealis mood.

- *(c)* Questions: Questions can express doubt, possibility, or wishes – all of which can signify the irrealis mood. For example: "Could you help me?" implies a potential action rather than a definite one.

- *(a)* Words enclosed in quotes: Sometimes words in quotes are used to refer to a hypothetical situation or a desired outcome. For instance, "If only I could 'fly'..." expresses a longing, reflecting an irrealis mood.

- Key Points About Irrealis Moods

- Function: They deal with situations that are possible, potential, hypothetical, wished for, or contrary to fact. *desire for*

- Contrast: The irrealis mood contrasts with the realis mood, which expresses facts or definite occurrences.

- Language: The way irrealis moods are expressed varies across different languages, with specific verb forms or auxiliary verbs often used.

(punctuation) *Grammar*

Q2

w: word
c: rating — (class)

□ Let P(w|c) represent the probability of a word given its rating. Further, let P(w) represent the probability of the word in the overall corpus. While analyzing the polarity of a word in a given corpus, what is the intuition behind dividing P(w|c) by P(w)?

Polarity analysis

$$\frac{P(w|c)}{P(w)} ?$$

a. To make it comparable across different ratings.

(Ans.)

b. To make it comparable across different words.

c. Both (1) and (2).

d. None of the above.

Q2

- Answer: 2

Q2

- Intuition:
- The core purpose of dividing P(w|c) by P(w) is to normalize for the overall frequency of words in the corpus and isolate the specific association of a word with a rating. Let's break down why:
- Word Frequency Bias: Some words are simply more common than others, regardless of sentiment. For example, "the" or "and" might appear frequently in both positive and negative reviews. Simply looking at P(w|c) wouldn't tell us if a word is a strong indicator of sentiment.
- Normalization Through P(w): Dividing by P(w) removes the influence of a word's general popularity. Let's consider two words:
- "Amazing": Appears often in positive reviews and is relatively common overall.
- "Terrible": Appears often in negative reviews but is less frequent in the entire corpus.
- Making Comparison Meaningful: The ratio of P(w|c) / P(w) helps us compare these words:
- "Amazing": Even with its overall frequency, it might still have a high ratio, indicating its strong association with positive sentiment.
- "Terrible": This would likely have an even higher ratio because despite being less frequent overall, it's strongly associated with negative reviews.
- Key Takeaway: This division lets you see how much more likely a word is to appear within a specific rating compared to its overall usage, revealing which words are truly significant sentiment indicators regardless of how often they are used generally

Q2

- The Problem with Raw Probability (P(w|c))
- Imagine we're analyzing sentiment and have two words: "amazing" and "terrible."
- Let's say P(amazing|positive) is high (meaning "amazing" often appears in positive reviews).
- Similarly, let P(terrible|negative) is high ("terrible" shows up frequently in negative reviews).
- On the surface, this might seem like both words are strong indicators of sentiment. But there's a potential issue:
- What if "amazing" is just a very common word in general (high P(amazing))?
- Even if it appears often in positive reviews (high P(amazing|positive)), its overall usage might skew the results.
- Normalization with P(w)
- Dividing by P(w) addresses this issue. P(w) represents the overall probability of the word in the entire corpus (collection of text). Here's what it accomplishes:
- Controls for Word Frequency: By dividing by P(w), we essentially remove the influence of a word's general usage.
- Focus on Relative Importance: We shift the focus to how much more likely "amazing" is to be in a positive review compared to its overall usage.
- Making Words Comparable
- Imagie after the division:
- The value for P(amazing|positive) / P(amazing) is significantly higher than P(terrible|negative) / P(terrible).
- This indicates that "amazing" is a much stronger indicator of a positive sentiment relative to its overall usage. Even if "terrible" has a high association with negative reviews, its frequent general use might dilute its significance.
- Therefore, dividing by P(w) allows you to compare words on a more even footing, highlighting which words are truly more indicative of a specific sentiment based on their usage within that context, rather than just their overall frequency.

Q3

semantic + grammatical words in language.

Which of the following lexicons are appropriate for valence?

a. Aroused, wide-awake

b. dominant, autonomous

c. happy, satisfied

d. stimulated, excited

Valence: the extent to which a word is positive or negative.

(emotional states)

Q3

- Answer: 3
- Valence: Valence refers to the inherent positivity or negativity of an emotion or experience. "Happy" and "satisfied" directly relate to positive valence, with varying intensities.
- Other options:
- Aroused, wide-awake: These describe activation levels or arousal, which is a different dimension of emotion than valence.
- Dominant, autonomous: These relate to feelings of power and control, not inherently tied to positive or negative valence.
- Stimulated, excited: While these can have positive connotations, they don't exclusively represent positive valence. Excitement can also be associated with nervousness or anxiety.
- Important Note: Sentiment analysis often incorporates multiple dimensions, including valence, arousal, and sometimes dominance. However, when specifically focusing on valence, words directly expressing positive and negative emotional states are the most relevant.

Q4

□ Which of the following is/are false?

a. The words 'bad', 'problem' represents negative emotion

b. The words 'love', 'sweet' belong to the class of affective processes

c. The words 'perhaps', 'guess' don't belong to cognitive processes

d. 'relaxation' is an example of low arousal, high pleasure word

Q4

- Answer: 3

- Solution: 'perhaps', 'guess' they are under cognitive processes.

cognitive

affective

→ intellect

interests

thoughts.

attitude

feelings

Q4

- The words 'bad', 'problem' represent negative emotion - TRUE These words generally carry negative connotations and associations.

- The words 'love', 'sweet' belong to the class of affective processes - TRUE Affective processes involve feelings and emotions, and 'love' and 'sweet' evoke positive feelings.

- The words 'perhaps', 'guess' don't belong to cognitive processes - FALSE 'Perhaps' and 'guess' indicate thought processes involving possibility, estimation, and uncertainty, which are cognitive activities.

- 'relaxation' is an example of low arousal, high pleasure word - TRUE Relaxation implies a calm, low-energy state combined with a sense of contentment.

- Summary

- The only false statement is: The words 'perhaps', 'guess' don't belong to cognitive processes

Q5

Which of the following is/are correct about Turney Algorithm

a. It uses Pointwise Mutual Information measure

b. Jaccard Coefficient also an be used to measure the co-occurence

c. This is used for phrase polarity task

d. It learns polarity of only a subset of phrase in the corpus

Q5

- Answer: 1

- Solution: Jaccard coefficient is the correct metric for co-occurence measure in Turney algo, and this algo learns the polarity of each phrase

Q5

$$PMI(w_1, w_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1) \cdot P(w_2)}$$

** (Most apt)

- It uses Pointwise Mutual Information measure - TRUE

- Core Mechanism: The Turney Algorithm fundamentally relies on Pointwise Mutual Information (PMI) to calculate the semantic orientation (positive or negative) of phrases. PMI is used to determine how much more likely two words or phrases are to occur together than would be expected by chance. This association with reference words ("excellent," "poor") is key to how it infers polarity.

- Why the other options are less clear-cut:

- Jaccard Coefficient also can be used to measure co-occurrence - POTENTIALLY TRUE   (modified Turney)

- While the Jaccard Coefficient could be used to identify phrases that frequently occur together within a corpus, it's not the primary method used by the Turney Algorithm.

- This is used for phrase polarity task - GENERALLY TRUE

- The main goal of the Turney Algorithm is indeed determining phrase polarity.

- It learns the polarity of only a subset of phrases in the corpus - MOSTLY TRUE

- The Turney Algorithm focuses on phrases containing adjectives or adverbs. However, one could argue that it indirectly infers polarity for a broader set of phrases due to the way PMI-based scoring works.

- Conclusion

- The safest and most accurate answer is indeed only option 1. The other options, while not strictly false, involve some degree of flexibility in interpretation or potential modifications to the original algorithm.
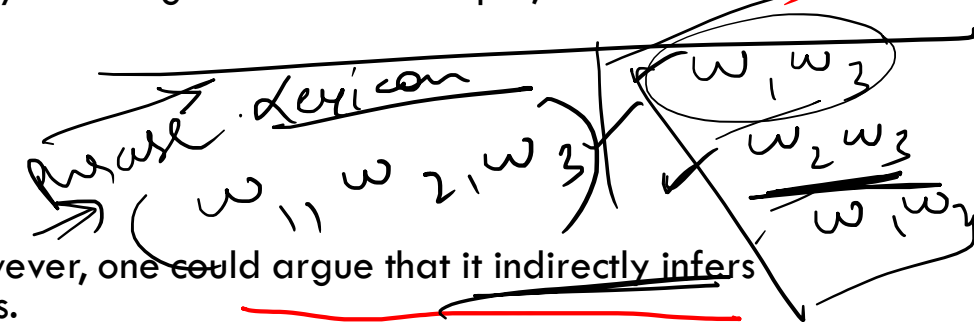
Q5

- Why the other options are less certain:

- Jaccard Coefficient also can be used to measure the co-occurrence - While technically true, the Turney Algorithm specifically relies on the Pointwise Mutual Information (PMI) measure for its calculations. Using the Jaccard Coefficient would represent a modification of the original method.

- This is used for phrase polarity task - This is generally true, as that's the main focus of Turney's Algorithm.

- It learns the polarity of only a subset of phrases in the corpus - While largely accurate, it could be argued that Turney's Algorithm indirectly infers the polarity of many phrases through its use of PMI-based scoring. The distinction here depends on how you interpret "learns".

- Therefore, the safest answer is indeed option 1, as it's the most unambiguously correct statement about the Turney Algorithm.

Q6

☐ Sentiment lexicons can be learned using intuitions such as:

a. Same polarity words are connected by "and"

b. Opposite polarity words are connected by "but"

c. Both (1) and (2)

d. None of the above

Q6

- The answer is Both (1) and (2). Here's why:

- Intuition 1: Same polarity words are connected by "and"

- Example: "The movie was exciting and thrilling."

- Logic: Words connected by "and" often indicate a continuation of the same sentiment. Here, both "exciting" and "thrilling" carry a positive connotation.

- Intuition 2: Opposite polarity words are connected by "but"

- Example: "The food was tasty but expensive."

- Logic: "But" often signals a contrast in sentiment. Here, "tasty" (positive) is contrasted with "expensive" (negative).

- How this helps Sentiment Lexicons:

- These intuitions form the basis for several semi-supervised and unsupervised approaches to constructing sentiment lexicons. Such methods work by:

- Seed Words: Starting with a small set of known "seed words" of positive and negative polarity.

- Pattern Identification: Analyzing large text corpora to identify how words are connected with "and," "but," and similar conjunctions.

- Expansion: Expanding the lexicon based on the identified patterns. For example, if both "good" and "great' frequently occur before "and," they likely carry similar polarity.

- Important Notes:

- These intuitions are not perfect rules and can have exceptions.

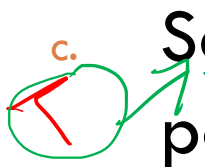- These techniques often work best in combination with other linguistic analysis methods.

Q7

[handwritten: [sentiment vs emotions] polarity shift tech. normalisation]

☐ Which of the following are true?

a. Emotions are instinctive and usually short-lived, whereas sentiments are more stable and valid for a longer period of time.

b. Polarity shift technique is performed to handle normalization problems

c. Sentiment lexicons can be learned using intuitions such as same polarity words are connected by "and.

d. 1 & 2

[handwritten: a,c b,c | a,b,c c,d c]

Q7

- Answer: 1, 3

Q7

- Emotions are instinctive and usually short-lived, whereas sentiments are more stable and valid for a longer period of time. - TRUE Emotions are often immediate reactions to stimuli, while sentiments reflect more enduring attitudes or beliefs.

- Polarity shift technique is performed to handle normalization problems - FALSE Polarity shift techniques address changes in sentiment expression, like sarcasm or negation, not normalization (making text forms consistent).

- Sentiment lexicons can be learned using intuitions such as same polarity words are connected by "and." - TRUE We discussed this earlier! Conjunctions like "and" and "but" provide clues to infer sentiment.

- Therefore, the correct answer is 1 & 3.

Handwritten annotations:
$P(w|c)$
$P(w)$

removing noise →
→ stop words
→ case nom.
→ lemmat'
→ Tokenisation

Q8

Consider the sentence: "The sound was cool; but, the network quality and screen were very dull". Which of the following are true?

a. Aspect: "sound", Sentiment: Positive, Opinion Phrase: "cool".

b. Aspect: "screen", Sentiment: Negative, Opinion Phrase: "very dull".

c. Aspect: "were", Sentiment: Negative, Opinion Phrase: "very dull".

d. Only (1) and (3)

[handwritten annotations:]

Amp (#)

e) Aspect: "n/w quality"; Sentiment: (-ve); Opinion: "very dull"

Stop word (grammatical structure)

Q8

- Answer: 1,2
- Solution: 'were' is not a aspect term

Q8

- Here's the correct analysis of the sentence:
- Aspect: "sound", Sentiment: Positive, Opinion Phrase: "cool". - TRUE This identifies the correct aspect and associated sentiment.
- Aspect: "screen", Sentiment: Negative, Opinion Phrase: "very dull". - TRUE This also accurately identifies an aspect and its sentiment.
- Aspect: "were", Sentiment: Negative, Opinion Phrase: "very dull". - FALSE Here's why:
- "Were" as a Verb: "Were" is a verb and not an aspect. Aspects refer to features or qualities of the subject (in this case, likely a phone or another device).
- Ambiguous Sentiment: While the overall sentiment about the network quality and screen is negative, "were" itself doesn't convey that sentiment.
- Therefore, the correct answer is Only (1) and (2)

Q9

☐ Which of the following is false?

a. Extraction of aspect-sentiment-opinion triplets is an easy task

b. Identification of Implicit aspects is challenging

c. Detecting same opinion phrases having different sentiments for different aspects is a challenging task

d. Aspect-based sentiment analysis problem can be solved by deep learning framework

Q9

- The false statement is: (a)

- Extraction of aspect-sentiment-opinion triplets is an easy task

- Here's why the other statements are considered true:

- Identification of Implicit aspects is challenging: Implicit aspects are not explicitly stated in the text and require inference or understanding of context, making their identification difficult.

- Detecting same opinion phrases having different sentiments for different aspects is a challenging task: This requires understanding nuances and how sentiment can shift depending on the specific aspect being discussed. For example, "long battery life" is positive for a phone but might be negative for a smartwatch.

- Aspect-based sentiment analysis problems can be solved by deep learning frameworks: Deep learning models like neural networks excel at handling complex language patterns, making them well-suited for aspect-based sentiment analysis tasks.

- Why aspect-sentiment-opinion triplet extraction is NOT easy:

- Ambiguity: Language has inherent ambiguity, and the same word or phrase might have different meanings or sentiments in various contexts.

- Relationships: Identifying the correct relationships between aspects, opinions, and sentiments can be difficult, especially with complex sentence structures.

- Data: Requires large amounts of labeled data for training accurate models, and acquiring this data can be time-consuming or expensive.

Q10

□ Consider the sentence: "The environment was great, however rent was very costly". Which of the following is/are true?

a. Aspect: "environment", Sentiment: Positive, Opinion Phrase: "great".

b. Aspect: "rent", Sentiment: Negative, Opinion Phrase: "very costly".

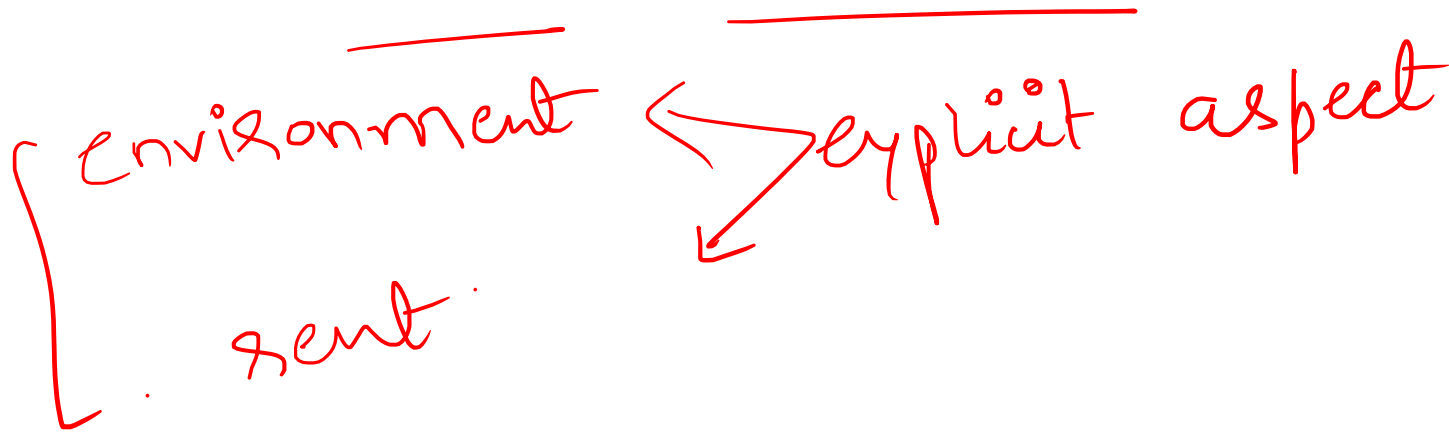c. Aspect: "price", Sentiment: Negative, Opinion Phrase: "very costly".

d. Only (a) and (b)

Imp

← cost

Note: consider implicit concepts also.

Q10

Answer: 1, 2, 3

☐ Solution: "price" is an implicit aspect.

environment ← explicit aspect
rent.

Q10

- Here's the reasoning behind why all three options could be considered correct:
- 1. Aspect: "environment", Sentiment: Positive, Opinion Phrase: "great".
- Unquestionably Correct: The word "environment" is the explicitly mentioned aspect, and "great" is a clear positive sentiment indicator for it.
- 2. Aspect: "rent", Sentiment: Negative, Opinion Phrase: "very costly".
- Also Unquestionably Correct: The word "rent" is specifically stated as the aspect, and "very costly" expresses a negative sentiment.
- 3. Aspect: "price", Sentiment: Negative, Opinion Phrase: "very costly".
- Implicit Aspect: "Price" is not explicitly mentioned, but "rent" is a type of price. This makes "price" a reasonable implicit aspect.
- General Sentiment: The negative sentiment expressed by "very costly" applies broadly to the concept of cost, even if the specific word "price" isn't there.

Q10

- (a) Correct:
- Aspect: "environment" - Clearly stated in the sentence.
- Sentiment: Positive - The word "great" is a positive sentiment indicator.
- Opinion Phrase: "great" - The word that modifies the aspect and conveys the sentiment.
- (b) Correct:
- Aspect: "rent" - Explicitly mentioned in the sentence.
- Sentiment: Negative - The phrase "very costly" conveys a negative feeling.
- Opinion Phrase: "very costly" - The words expressing the sentiment about the aspect.

Q10

- Why option (c) could be considered acceptable:
- Implicit Aspect: While "price" isn't explicitly stated, one could argue that "rent" is a specific type of price, making "price" a valid implicit aspect.
- General Sentiment: The sentiment "negative" broadly applies to the concept of cost, even if the specific word "price" isn't used.
- Interpretation: Some evaluations of aspect-based sentiment analysis might allow for slightly more flexibility in interpreting implicit aspects and general sentiment.
- However, here's why the strictest interpretation favors only (a) and (b):
- Specificity: Aspect-based sentiment analysis often emphasizes precision in identifying the exact terms used in the text. In this case, "rent" is the explicitly stated aspect.
- Nuance: While related, "rent" and "price" have different connotations. Rent specifically refers to recurring payment for temporary use, which might carry a different sentiment than the broader concept of price.
- Conclusion:
- Ultimately, whether option (c) is strictly correct depends on the specific evaluation criteria being used. There's a case to be made for both interpretations.
- It highlights the importance of understanding the nuances of aspect-based sentiment analysis and potential variations in how tasks might be evaluated

Q11

The word nervous and jealous falls under which affective state typology?

a. Emotion

b. Mood

c. Personality traits

d. Interpersonal stances

Q11

- Ans: a
- The words "nervous" and "jealous" fall under the category of Emotion. Here's why:
- Emotions: Brief, intense, and focused reactions to specific events, objects, or situations. Nervousness and jealousy fit this definition well.
- Moods: Lower intensity, more diffuse, and longer-lasting than emotions. Examples include feeling cheerful or gloomy.
- Personality Traits: Enduring patterns of thinking, feeling, and behaving. These are more stable than emotions or moods. Examples include extroversion or introversion.
- Interpersonal Stances: Positions we take towards others in social interactions. These often involve attitudes or beliefs. Examples include friendliness or dominance.
- Important Note: While moods can influence the type and intensity of emotions we experience, nervous and jealous clearly fit the category of direct emotional responses.

emotions → moods

Duration:

emotions << Moods
emotions << sentiment

Q12

Which of the following is/are tokenization issues for Sentiment Analysis?

a. Handling emotions

b. Handling negations

c. Word lengthening

d. None

Q12

Solution:

B,c

Q12
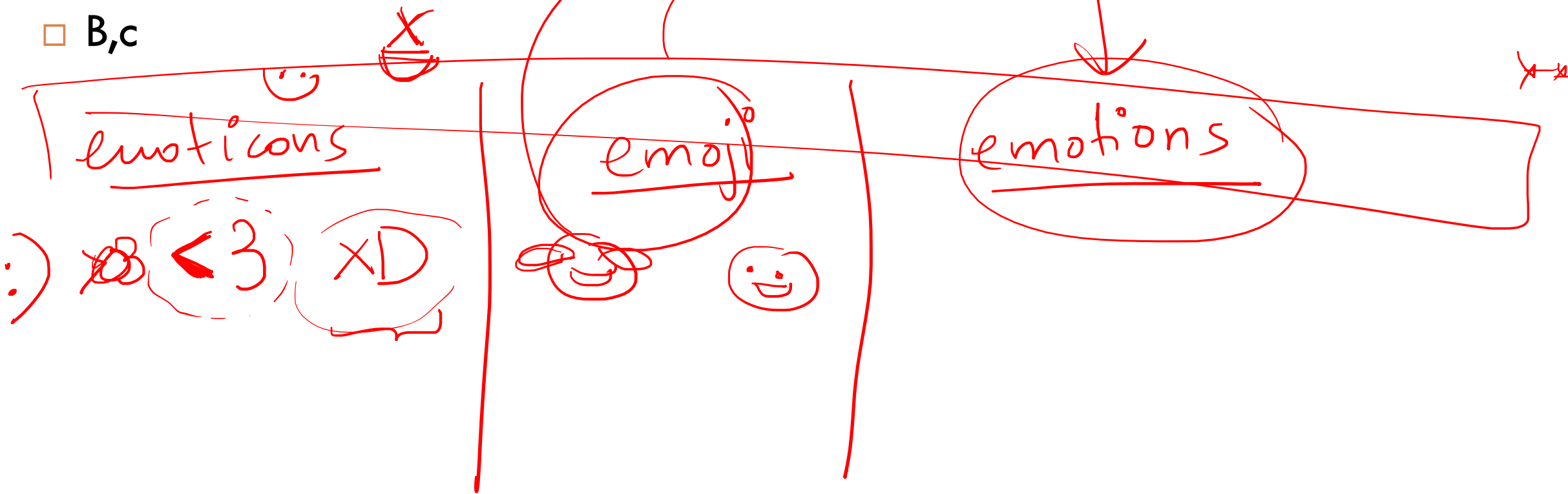
[handwritten: "didn't" → "did not"]

- Handling emotions: While dealing with emotions is crucial in sentiment analysis, it's not directly a tokenization issue. Tokenization is about how you break down text into units (words, phrases, etc.).

- Handling negations: Absolutely! Negations (e.g., "not", "didn't") flip sentiment and need careful handling. Incorrectly tokenizing negation words with the word they modify leads to misinterpretations.

- Word lengthening: This is a significant tokenization challenge. Informal words with lengthened characters (e.g., "loooove", "yaaaaas") can change the emphasis or meaning, so accurate tokenization is crucial for sentiment interpretation.

- Therefore, the correct answer is: Handling negations and Word lengthening

- Here's why tokenization matters for sentiment analysis:

- Accurate Representation: Sentiment is often tied to specific words and word combinations. Tokenization errors can obscure or break these sentiment indicators.

- Context Preservation: Understanding how negation words modify surrounding words is only possible with proper tokenization

[handwritten: luv, I ❤️ u, loove]

Q13

Which of the following is/are false?

a. The picture quality of the camera is great- 'picture quality' is an implicit aspect

b. This laptop is expensive- 'expensive' is an implicit aspect expression

c. This phone will not fit in a pocket- 'fit in a pocket' is an explicit aspect expression indicates the aspect size

d. All of the above

Handwritten annotations:

implicit /explicit aspects

F

F

f

(implicit)

This laptop is expensive → expensive: explicit price/cost: implicit

Q13

- Ans: d
- The picture quality of the camera is great- 'picture quality' is an implicit aspect - FALSE. "Picture quality" is the explicitly stated aspect in this sentence.
- This laptop is expensive- 'expensive' is an implicit aspect expression - TRUE. While we understand the sentiment is about the price, the aspect (price) is not explicitly mentioned.
- This phone will not fit in a pocket- 'fit in a pocket' is an explicit aspect expression and indicates the aspect size - FALSE. While "fit in a pocket" is explicit, it speaks to the aspect of portability or convenience rather than the size of the phone itself.
- Therefore, the correct answer is All of the above.

Q14

Consider the sentence: "The display was cool' but the network quality and camera were very dull." Which of the following are true?

a. Aspect: "display", Sentiment: Positive", Opinion phrase: "cool"

b. Aspect: "camera", Sentiment: Negative", Opinion phrase: "very dull"

c. Aspect: "were", Sentiment: Negative", Opinion phrase: "very dull"

d. Only 1 and 3

Ans: a, b

This laptop is small, heavy ⇒ heavy: explicit aspect
weight: implicit

Q14

- The correct answer is actually Only 1 and 2. Here's the breakdown:

- 1. Aspect: "display", Sentiment: Positive", Opinion phrase: "cool" - CORRECT. This clearly identifies the aspect and corresponding sentiment.

- 2. Aspect: "camera", Sentiment: Negative", Opinion phrase: "very dull" - CORRECT. This also correctly identifies the aspect, sentiment, and the opinion phrase expressing that sentiment.

- 3. Aspect: "were", Sentiment: Negative", Opinion phrase: "very dull" - INCORRECT. Here's why:

- "Were" is a verb and doesn't represent an aspect (a feature or attribute being discussed).

- While the sentiment "negative" is associated with the sentence overall, "were" itself doesn't directly express that sentiment.

- Key Points:

- Aspects are nouns or noun phrases: They refer to specific components, features, or attributes of the subject being discussed.

- Opinion phrases modify aspects: They express the actual sentiment using adjectives, adverbs, or similar terms

Q15

The word "loving" and "buoyant" falls under which affective state typology?

a. Attitudes, Mood

b. Emotion, Attitudes

c. Emotion, Mood

d. Mood, Personality traits

Q15

- The correct answer is Emotion, Mood. Here's why:
- Emotion: "Loving"
- Emotions are brief, intense reactions to specific stimuli. "Loving" often reflects an emotional response to a person, experience, or object.
- Mood: "Buoyant"
- Moods are longer-lasting, less intense, and more diffuse than emotions. "Buoyant" describes a positive, cheerful, and upbeat mood state.
- Why the other options are less suitable:
- Attitudes: Enduring beliefs or predispositions towards objects or people. While "loving" could relate to an attitude towards someone, it's primarily an emotional response.
- Personality Traits: Stable patterns of behavior, thought, and feeling. While "buoyant" might be exhibited by someone with a generally optimistic personality, it's primarily a description of a mood state.

Q16

- Which of the following is/are True about the Turney algorithm?

a. It extracts a phrasal lexicon from reviews

b. Learns polarity of each phrase

c. Rate a review by the average polarity of its phrases

d. All of the above

Q16

- The correct answer is All of the above. Here's a breakdown of why each statement is true:
- It extracts a phrasal lexicon from reviews: The Turney algorithm heavily relies on identifying phrases consisting of adjectives or adverbs followed by nouns, which form the basis of its lexicon.
- Learns polarity of each phrase: It calculates the semantic orientation (positive or negative) of phrases using Pointwise Mutual Information (PMI) and the proximity of these phrases to reference words like "excellent" or "poor."
- Rate a review by the average polarity of its phrases: After identifying the polarity of phrases within a review, the algorithm averages these polarities to determine an estimated overall sentiment for the review.
- Key Points:
- Unsupervised Approach: The Turney algorithm is unsupervised, meaning it doesn't require pre-labeled training data. It learns from patterns within a large text corpus.
- Limitations: It has its limitations in accuracy and handling complex linguistic nuances.

Contact

- **Email**: phd2001201007@iiti.ac.in

- **Youtube Channel:**
  https://www.youtube.com/channel/UCbyhbe4RvYYbp_IMnYVSfzw

- **Lecture Slides:**
  https://drive.google.com/drive/folders/17vSA5Qjnr7ghIxNliC36Pn5F
  Fok_Kvb8?usp=sharing