

1. Define the following with suitable examples:

- Opinion
- Direct Opinion
- Comparative Opinion
- Implicit And Explicit Opinion
- Opinion words and phrases

ANSWER: —————

Opinions are subjective, influenced by people's personal experiences and beliefs, thus, can vary from person to person. Example: "I love comfortable fashion. T-shirts and jeans are my go to outfits."

Direct opinion is an opinion given by an individual which could be positive or negative about an entity by not comparing it with any other similar entity. Example: "This book has so much suspense and thrillers."

Comparative opinion is an opinion given by an individual which could be positive or negative about an entity by comparing it with any other similar entity. Example: "This book has so much suspense and thrillers than the last book I read."

Explicit opinion is an opinion which is clear and direct regarding any object, giving clear and straightforward message. Whereas, implicit opinion are not clearly indicated but rather implied through a context or a tone. Example "We found the flight to be exceptionally comfortable" is explicit and "The flight we experienced offered a level of comfort that met our expectations." is implicit.

Opinion words are terms which are frequently used to convey either positive or negative opinions/feelings. Opinion phrases are phrases or idioms that are used to express positive or negative feelings. Example "Good" is opinion word and "Silver lining" is phrase.

[Reference: "1. The Problem of Sentiment Analysis"

https://www.researchgate.net/profile/Bing-Liu-120/publication/228667268_Sentiment_analysis_and_subjectivity/links/5472bbea0cf24bc8ea199f7c/Sentiment-analysis-and-subjectivity.pdf]

2. Discuss the task of subjectivity classification with examples.

ANSWER: —————

It's a task of determining whether a sentence has opinion or not. Suppose we have sentence, then first task is subjectivity classification and second is sentence-level sentiment classification. We determine whether sentence is subjective or objective. Further identifying if opinion is positive or negative using sentence-level sentiment classification.

To achieve this, ML approaches are involved, but needs lots of manual work in labelling training set. However, this problem was resolved using bootstrapping approach, replacing manual effort by automatically labelling training set.

The algorithm begins with two high-precision classifiers, HP-Subj and HP-Obj, to identify subjective and objective sentences using strong subjectivity clues. They aim at achieving high precision, but overlook many sentences not fitting strict criteria, resulting in low recall. Extracted sentences are added to training data, and labelled with their subjectivity classification. New classifiers are built using patterns learned from previous iteration. Sentences are added in training set, and algorithm moves to following iteration.

Example “Smartphone has 64 mega pixel camera.” is objectivity and “I love sharp colours of photos from smartphone’s camera.” is subjectivity.

[**Reference:** “2.2 Sentence-Level Subjectivity and Sentiment Classification”

https://www.researchgate.net/profile/Bing-Liu-120/publication/228667268_Sentiment_analysis_and_subjectivity/links/5472bbea0cf24bc8ea199f7c/Sentiment-analysis-and-subjectivity.pdf]

3. What are the different levels of sentiment classification? Explain with suitable examples.

ANSWER: —————

1. **Document-level sentiment classification:** In this, whether the document exhibits overall positive or negative opinion is determined, assuming that document contains opinions on single product and opinion is from single person. Example, “I bought new car, it gives great mileage. It also has automatic emergency brake feature which I love.”, overall it gives positive opinion.
2. **Sentence-level sentiment classification:** In this, if the sentence is subjective, then it determines if it exhibits positive or negative opinion. Example, “I bought new car, it gives great mileage. But front and rear parking sensors does not perform great”. Here former sentence gives positive opinion, and later one gives negative opinion.
3. **Opinion lexicon generation:** In this, whether the opinion of a document is positive or negative is based on the individual documents features and characteristics. Example, “I bought new car, it gives great mileage. But front and rear parking sensors does not perform great”, here “mileage” is positive, but “parking sensors” is negative.

[**Reference:** “2.1 Document-Level Sentiment Classification”, “2.2 Sentence-Level Subjectivity and Sentiment Classification”, “2.3 Opinion Lexicon Generation”

https://www.researchgate.net/profile/Bing-Liu-120/publication/228667268_Sentiment_analysis_and_subjectivity/links/5472bbea0cf24bc8ea199f7c/Sentiment-analysis-and-subjectivity.pdf]

4. Discuss the existing techniques for document-level sentiment classification.

ANSWER: —————

There are two primary categories of the techniques for document-level sentiment classification:

1. Classification based on Supervised Learning
2. Classification based on Unsupervised learning

Classification based on Supervised Learning

Positive and negative labels serve as the two main categories for supervised learning problems that deal with sentiment classification. Star ratings are used as labelled data, where high-star reviews denotes positive opinion and low-star denotes negative opinion. The significant difference between sentiment classification and topic-based text classification is that topic-related terms are crucial in topic-based classification whereas, words that elicit strong emotions, like "great," "excellent," "horrible," and "bad," are crucial. Support vector machines (SVM) and naive Bayesian techniques have both been effectively used in Sentiment classification.

Classification based on Unsupervised learning

Unsupervised learning is an obvious choice for this job since sentiment analysis mainly relies on words and phrases that are associated with opinions. A particular approach describes a procedure with three steps:

STEP 1 - Extracting word combinations that contain adjectives or adverbs is the first stage. The algorithm finds pairs of words that are sequential and contain an adjective or adverb and a context word. This combination is chosen if it follows specific patterns based on their part-of-speech markers.

STEP 2 - At this stage, the algorithm analyses the retrieved word combinations' sentiment orientation using a metric known as pointwise mutual information (PMI), which is specified in below equation.

$$PMI(term_1, term_2) = \log_2(Pr(term_1 \wedge term_2)/Pr(term_1)Pr(term_2))$$

STEP 3 - The programme determines the average sentiment orientation for all phrases taken out of the review. The review is categorised as recommended if the average is positive, otherwise, it is categorised as not recommended.

[Reference: “2.1.1 Classification Based on Supervised Learning”, “2.1.2 Classification Based on Unsupervised Learning”

https://www.researchgate.net/profile/Bing-Liu-120/publication/228667268_Sentiment_analysis_and_subjectivity/links/5472bbea0cf24bc8ea199f7c/Sentiment-analysis-and-subjectivity.pdf]

5. Discuss the shortcomings of document-level sentiment analysis and sentence-level sentiment classification.

ANSWER: —————

Document-level sentiment analysis shortcomings:

1. Challenging to categorise longer documents into single category due to complex opinions.
2. Various subjects may be covered in certain documents, leading inaccuracy to calculate a single sentiment score.
3. Difficult to capture sentiment of some statements at document level since it can vary depending on larger context of the document.
4. Analysis at the document level offers a general sentiment but may overlook variations in the material.

Sentence-level sentiment analysis shortcomings:

1. It might be difficult to tell if a statement lacks sentiment-bearing words or is indeed neutral.
2. It might be difficult to categorise sentences containing several clauses or contradictory emotions, such as "I love the city but hate the weather."
3. These language patterns can result in categorization errors at the sentence level and are extremely hard to spot.
4. Despite being more specific than texts, phrases sometimes lack enough information on their own. In one context, a phrase could be positive but in another, it might be negative.

[**Reference:** "2.1 Document-Level Sentiment Classification", "2.2 Sentence-Level Subjectivity and Sentiment Classification"]

https://www.researchgate.net/profile/Bing-Liu-120/publication/228667268_Sentiment_analysis_and_subjectivity/links/5472bbea0cf24bc8ea199f7c/Sentiment-analysis-and-subjectivity.pdf

6. Explain with examples how and where the linguistic concepts of i. Part of speech tagging and ii. syntactic dependency parsing are used in document-level sentiment classification.

ANSWER: —————

Part of speech (POS) tagging

POS tagging identifies the appropriate part of speech for each word in a document based on its meaning and context. Nouns (N), verbs (V), adjectives (ADJ), and adverbs (ADV) are examples of

common tags. Example word "like". "I like the taste of chocolate" is a favourable feeling when used as a verb. "The dance like that" as a preposition lacks feeling.

1. Incorporated into machine learning models as features. For instance, a sentence may be sentiment-bearing if adjectives and adverbs are used frequently.
2. Adjectives and Adverbs, since these words frequently express emotion, they are particularly crucial in sentiment analysis. For instance, the adverb "exceptionally" and the adjective "captivating" convey the feeling in the statement "The book was exceptionally captivating."
3. Depending on their position in a phrase, words might have many meanings. We can better understand the sentiment if we can pinpoint the proper POS.

Syntactic Dependency Parsing

Dependency parsing establishes linkages between "head" words and words that modify or depend on them to ascertain grammatical structure of a phrase. Example "I liked the design but not the colour", can be determined via dependency parsing that the positive sentiment is directed at the "design" and the negative sentiment is directed at the "colour".

1. Dependency parsing can be used to determine the entities that an emotion is directed towards. This is especially important when discussing several entities in documents.
2. Dependency parsers are capable of spotting negation structures, which can change the meaning of a statement. For instance, the negative "not" modifies "interesting" in the sentence "The movie is not interesting."
3. Dependency parsing may be used to identify the sentiments in phrases that contain several feelings that are connected to various targets.

[**Reference:** "2.1 Document-Level Sentiment Classification"

https://www.researchgate.net/profile/Bing-Liu-120/publication/228667268_Sentiment_analysis_and_subjectivity/links/5472bbea0cf24bc8ea199f7c/Sentiment-analysis-and-subjectivity.pdf]

7. What is pointwise mutual information (PMI)? Explain how it is used for opinion classification with a method based on Unsupervised learning.

ANSWER: —————

PMI is a metric used in information theory and statistics. It contrasts the likelihood of two occurrences happening simultaneously with the likelihood of the same events occurring independently. Following formula to calculate association between two words,

$$PMI(term_1, term_2) = \log_2(Pr(term_1 \wedge term_2)/Pr(term_1)Pr(term_2))$$

Where,

$PMI(term_1, term_2)$ = Pointwise Mutual Information between two terms

$Pr(term_1 \wedge term_2)$ = Probability of $term_1$ and $term_2$ occurring together

$Pr(term_1)$ = Probability of $term_1$ occurring

$Pr(term_2)$ = Probability of $term_2$ occurring

PMI provides a helpful method for developing a sentiment vocabulary without the requirement for supervision.

When calculating PMI, a target word's emotion may be determined by contrasting it with both accepted positive and negative phrases.

PMI has ability to assess phrase sentiment. Knows the difference between “great” and “not great”.

[**Reference:** “2.1.2 Classification Based on Unsupervised Learning”

https://www.researchgate.net/profile/Bing-Liu-120/publication/228667268_Sentiment_analysis_and_subjectivity/links/5472bbea0cf24bc8ea199f7c/Sentiment-analysis-and-subjectivity.pdf]

8. Discuss three applications of sentiment classification.

ANSWER: — — — — —

1. **Social Media Monitoring:** Great amount of data is available of people reacting to every little thing online. Sentiment classification can help understanding the trends people are into and gather information about their likes and dislikes on different products and topics, which can also help business to understand their target audience.
2. **Brand Mentions:** Sentiment classification can help the brands in understanding the areas where they are good and bad, and also help understanding their audience which can lead to improving their brand decisions and reputation.
3. **Understanding Employee:** Using sentiment classification, companies can understand the sentiments of their employees like how satisfied they are, or their behaviour in certain scenarios, which can lead to improved productivity and satisfied employees.