Name: Vikrant Singh Jamwal
Student ID: 23104534

Sentiment Analysis and Subjectivity

Solution 1: Defining terms with suitable examples:

1. **Opinion**
   An opinion is a subjective expression of someone's belief or feeling on a particular topic. It can be positive, negative or neutral.

   Example:
   "This is an interesting Assignment"- Positive.
   "Assignment seems boring"- Negative.
   "It is time taking, but educative"- Neutral.

2. **Direct Opinion**
   Direct opinion refers to statements that expresses a positive or negative opinion directly without any comparison with other objects.

   Example:
   "Galway city is very beautiful" – Positive Direct opinion on Galway.
   "The food here is bad" – Negative Direct opinion on food.

3. **Comparative Opinion**
   Comparative opinion is an expression that compares one object to other objects and provides a preference or a judgement.

   Example:
   "NLP course is more interesting than other courses" – Comparing NLP and other courses and giving a judgment on which seems more interesting.
   "My phone has a better camera than my friend's phone" – Comparing and giving positive opinion to camera of my phone to my friend's phone.

4. **Implicit Opinion and Explicit opinion**
   Implicit opinion does not clearly or explicitly express whether the statement is positive or negative. On the other hand, Explicit opinion clearly states the expression as positive or negative.

   Example:
   "I love this book" – Explicit opinion.
   "This car is slow" – Implicit opinion - A negative opinion without explicitly stating it.

5. **Opinion words and phrases**
   Words and phrases that indicates sentiment in a statement are opinion words or opinion phrases.

   Example:
   Words- "beautiful", "awful", "fantastic", Phrases- "In my opinion", "According to me"

Name: Vikrant Singh Jamwal
Student ID: 23104534

Solution 2: Task of subjectivity classification

Subjectivity classification is a task of determining whether a sentence is opiniated or not. As opinions are subjective and not objective. It aims to classify the expression as subjective or objective. Objective statements are factual statements and are not opinions.

Example:
"I believe we had a nice time working together" - Here phrase "I believe", indicates an opinion, making the expression a subjective expression.
"The sun rises from the west" - Here the statement is factual and no opinion is present, hence it is an objective expression.

Solution 3: Different Levels of Sentiment classification

There are two levels of sentiment analysis:

1) **Document-Level Sentiment Classification**
   This level of sentiment classification aims at to determining the overall sentiment of the subjective/opiniated document on an object. In other words, determining the opinion, document expresses on an object. Most existing techniques are supervised learning classification techniques.

   Example:
   A document containing customer feedback of a product. Determining whether the opinion on the product (object) is positive or negative.

2) **Sentence Level Sentiment Classification**
   This level of sentiment classification aims at determining the opinion sentiment of a sentence on an object or its features. This level is important as it filters out non opiniated sentences present. It includes two sub steps. Subjective classification where sentence is determined as subjective or not and then sentiment classification if the subject is subjective. It can also be achieved by traditional supervised learning as this is also a classification problem.

   Example:
   A feedback of a single user could be of multiple sentences. Each sentence is checked if subjective or not and then each subjective sentence is stated as positive or negative.

Solution 4: Existing Techniques of Document Level sentiment classification

Existing techniques for document-level sentiment classification are:

Supervised learning, sentiment classification is formulated as a binary classification problem with positive and negative class labels. Providing weights to opinion terms and phrases and predicting its sentiment. Negation is also taken into account as a positive opinion could be negated with a "not". Techniques such as naïve Bayesian and support vector machines (SVM) have been used with features like unigrams to achieve good performance.

Unsupervised methods, on the other hand, do not require labelled data but rely on patterns and statistical analysis to determine sentiment. We extract the adjectives or adverbs (as research tells that these words best determine an opinion or subjectivity) with their neighbour word and performs classification based on some fixed syntactic phrases that are likely to be used to express opinions. However, supervised learning approaches are more commonly used in sentiment classification.

Solution 5: Shortcomings of Document-level sentiment analysis and Sentence level sentiment classification

Document-level sentiment classification, does not provide detailed information about specific aspects or features of the document that are liked or disliked by the author. Another limitation is that it assumes that the document expresses opinions on a single object and the opinions are from a single opinion holder. However, there can be multiple objects and multiple complex and comparative opinions. Document level sentiment classification cannot be used in such situations for getting an accurate information from the document.

Sentence-level sentiment classification assumes that each sentence will express a single opinion, hence is not suitable for compound sentences. It was pointed out that not only a single sentence may contain multiple opinions, but also both subjective and factual clauses. It is useful to pinpoint such clauses. It is also important to identify the strength of opinions which sentence level sentiment classification fails at.

Solution 6: Use of linguistic concepts in Document-Level Classification

1) **Part of Speech (POS tag)**
   Part of speech tagging is used in document-level sentiment classification to identify the grammatical category of each word in a document. It was found in research that adverbs and adjectives are important indicators of subjectivities and opinions. Thus, adjectives and adverbs have been treated as special features, making part of speech an important aspect in document level sentiment classification.
   Example:
   Adjectives and adverbs like “great”, “amazing” indicates positive sentiment, while “horrible” suggests a negative sentiment.

2) **Syntactic Dependency**
   Syntactic dependency parsing is used in document-level sentiment classification to analyse the relationships between words in a sentence and understand the syntactic structure of the document accurately. Understanding the relationship between words also helps in determining negation accurately.

   Example:
   “The book was amazing” and “The book was not bad”, although adjectives and adverbs used are of opposite sentiments but through syntactic dependency we can figure out that both sentiments are positive as negation is used.

Name: Vikrant Singh Jamwal
Student ID: 23104534

Solution 7: Pointwise mutual information (PMI)

Pointwise mutual information is measure of association of co-occurring words. It can be calculated by taking log of division of the probability of the words occurring together by the product of the probabilities of both occurring individually.

$$PMI = \log\{prob(word1 \wedge word2) / (prob(word1) * prob(word2))\}$$

Here, prob(word1 ^ word2) is the co-occurrence probability of two words, and prob(word1) and prob(word2) gives the probability that the two terms co-occur if they are statistically independent. The ratio between is thus a measure of the degree of statistical dependence between them. The log of this ratio is the amount of information that we acquire about the presence of one of the words when we observe the other.

As for **unsupervised learning**, classification is done by some fixed syntactic phrases that are likely to be used to express opinions. For comparing these fixed syntactic phrases with the test document, we use bigrams of words and analyse them by finding relation and similarity between words. To calculate their co-occurrence probability we use PMI. PMI provides information of opinionated words which most likely to occur with other words making it easy to predict their sentiment.

Solution 8: Applications of Sentiment Classification

1) **Entertainment Industry**

Sentiment classification Is applied in entertainment industry to get the reviews of what audience liked or disliked and what is the current trend or mindset of the society to create content accordingly. Example:
Movie reviews are used to understand the sentiments of the audience to analyse what are the mistakes and what could be the solution to those in future.

2) **Healthcare**

Applying sentiment classification to healthcare institutions will help to understand what could be improved to uplift patients satisfaction level and trust.
Example:
Long queues at emergency department can be risky for emergency situations and understanding patients sentiments will help improve management.

3) **Consumer Feedbacks**

Mostly companies ask for a feedback or reviews for the product customers bought. Using sentiment analysis will help understand the customer satisfaction level.
Example:
Consistent negative reviews from a location suggests need for some improvements to the delivery companies.