**INTRODUCTION TO SENTIMENT ANALYSIS**

**Introduction:**

Every individual before conducting a research or to make day-to-day decisions often o for a massive opinion .Today, opinions are being collected for each and every thing- right from purchasing a product till the serious decisions that impacts political or economic scenario of the nation.

**Definition:**

Sentiment Analysis is also referred to as subjectivity analysis or opinion mining or appraisal extraction. It is a process of identifying and classifying the opinions to asses the public perspect towards the given product or any object( in terms of Positive, Negative or Neutral).

**Terminology in Sentiment Analysis:**

The various terms often used in Sentiment Analysis are:

* Opinion- is a statement which is yet not concluded.
* View- expresses the subjectivity of the opinion.
* Belief- indicates the intentions and rational views.
* Conviction- applicable to a strongly held belief.
* Persuasion- implies evidence to the belief.
* Sentiment-gives a grounded opinion of one’s own feelings.

Wiebe, a prominent Natural Language Processing (NLP) researcher, used Quirk’s definition of the private state when tracking point of view in narrative (Wiebe, 1994). She defines private state as a tuple

(p, experiencer, attitude, object)

**Sentiment Analysis:**

In Sentiment Analysis subjectivity of the opinion is studied which is expressed in linguistic units. Such as words, phrases or sentence. To be more it is suggested to use small linguistic units.

Generally, Opinion appears in two contexts –explicit(Accurate and Specific views) and implicit(Fuzzy and imprecise views).Researchers so far focused only on explicit opinions which can be easily assessed.

One of the key feature of the present context is sentiment polarity which is usually takes positive and negative .Another important aspect is target which is an object or a concept or a person or anything.

**Example of Sentiment Research:**

OpSec Security provides consumer feedback data analysis to their customers in order to help them in identifying the needs of market and target customer and their market position.

**Goals of Sentiment Analysis:**

Sentiment analysis invokes the following tasks.

* **Opinion Sentiment Detection:** Classifying the objective or subjective based on the adjective used in opinion.
* **Polarity Classification:** Classifying the opinions on two opposite sentiment polarity using multipoint scale.
* **Discovery of Opinion target**
* **Feature Extraction:** Defining the various features or components or attribute of an object.

**Methodologies:**

1. **Classification:**

Machine learning and part-of speech taggin are most common tools used for classification.

* Term Presence Vs Frequency
* n-grams
* Part-of-Speech
* Syntax
* Negations

1. **Identifying the semantic orientation of words:**

* **Lexicon:** Lexicon is the Simplest method with binary classification of words into another positive and negative or objective and subjective. The strength of the label can be determined using the concept of Fuzzy lexicons.
* **Training Documents:** Training documents(Labelled-Manual or Unlabelled-Automatically) helps to classify the sentiment/opinion using statistical methods and machine learning tools.

1. **Identifying Semantic orientation of Sentiment Sentences and phrases:**

After identifying the semantic orientation of the words, It is usually extended to phrase and sentence by categorizing the average repeated polarities of the words in the sentence.

1. **Semantic orientation of documents:**

It is not possible for large documents such as blocks but for small documents and mails-Document labeling is used.

One of the most popular, and simple, methods is a linear combination of allpolarities. For example, Dave et al. (Dave et al., 2003) and Turney et al. (Turney and Littman, 2003) use averaging to determine the polarity of documents. Thpolarity of the document can be expressed as

class(di) =

C; eval(di) > 0

C0; eval(di) < 0 (1)

eval(di) =∑score(tj) (2)

The team at University of Illinois at Chicago (Voorhees and Buckland, 2007) used a set of rules with thresholds to label the documents:

* Firstly, if both positive and negative opinions are strong in the document, the document should be mixed.
* Otherwise, if one type of the opinions is strong, the document is labeled to that type.
* Finally, if there are no strong opinions either way, the document is labeled as mixed.

1. **Feature Extraction:**

Target is the further more in sentiment analysis. If a company wants to know the specific likes/dislikes about the product they often go for feature extraction ,for which parts of speech tagging is used.

A common approach is to use the part-of-speech (POS) tags to construct templates of how sentiment is applied to objects. For example, Bing Liu et al. (Liu et al., 2005) use this process for a phrase “included memory is stingy”:

1. Perform part-of-speech (POS) tagging and remove digits:

“<V>included <N>memory <V>is <Adj>stingy”

2. Replace the actual feature words in a sentence with [feature]:

“<V>included <N>[feature] <V>is <Adj>stingy”

3. Use n-gram to produce shorter segments from long ones:

“<Adj>included <N>[feature] <V>is” “<N>[feature] <V>is <Adj>stingy”

4. Distinguish duplicate tags by giving them numbers:

“<Adj1>included <N1>[feature] <V1>is”

5. Perform word stemming

Table 1: Feature Information

|  |  |
| --- | --- |
| **Explicit Features Examples** | **Properties Scanner Size** |
| Parts Scanner Cover  Related Concepts Scanner Image  Explicit Features Examples  Parts Scanner Cover  Related Concepts Scanner Image | Features of Parts Battery Life  Related Concept’s Features Scanner Image Size  Properties Scanner Size  Features of Parts Battery Life  Related Concept’s Features Scanner Image Size |

1. **Comparative Sentence Identification:**

It presents the relation between among the two or more objects based on the similar an difference .The comparative may be gradable(based on relation’s greater than/less than/equal).Usually data mining techniques are used to identify comparative sentences.



Figure 2: Visual comparison of consumer opinions on two products

**Conclusion:**

Though the sentiment analysis uses different tool and techniques to extract the opinions, The text requires still more advance approaches to deal with the two major problems

1. The present algorithms fail to extract in the more complex settings such as political forums and news articles.
2. As sentiment is topic specific, General lexicons should not be used for different objects.