Assignment 3

REPORT 1: SENTIMENT ANALYSIS

MEGHNA RAMACHANDRA HOLLA | B00812604

Sentiment Analysis

Every day millions of data is generated in textual format over the internet. Although understanding this huge amount of unstructured data is difficult, it is important to structure it and analyze it. Through these text documents, analyzing what a person feels towards a subject or an object gives rise to what is known as Text Analytics and Text mining. The process of analyzing text (Text Analytics) and extracting data (Text Mining) to understand opinions (Opinion Mining) made by people on the internet is Sentiment Analysis.

A. Text Mining, Text Analytics and Opinion Mining

Text Mining is the precursor of Text Analytics. Text mining is a process of extracting information from text and cleaning it. Text Analytics runs statistical algorithms on the extracted data to generate useful information. Opinion Mining is a natural language processing (NLP) task that categorizes text into positive, neutral or negative based on the opinionated content that was derived. Opinions can also be of two types – subjective and objective i.e., opinion on a subject like an issue or phenomenon and opinion on product, person, organization.

B. Details of Sentiment Analysis

Sentiment Analysis is an automated approach to understand opinions and sentiments from textual content (e.g. Twitter/Facebook data) and derive attributes from an expression in the text like Polarity (positive or negative opinion), Subject (the topic being addressed) and Opinion Holder (the person expressing the opinion). Sentiment analysis is useful for organizations to understand their customer needs and opinions. Based on the automated analysis of customer feedback, the organization will send recommendations to a customer about products or improve their product. Sentiment analysis can be applied to different levels of text like a paragraph, a sentence or even an expression within a sentence.

C. Methodology of Sentiment Analysis

1. Identifying the goal and purpose of using Sentiment Analysis to determine the keywords for analysis.
2. Prior to working on Sentiment analysis, our text needs to be cleaned. This can be done by the following processes:
   - Removing numbers, punctuation, URLs, Links
   - Removing stopwords (e.g. a, an, the, but, etc.)
   - Stemming words (reducing words to their root form)
   - Lemmatisation process (grouping similar or related words e.g. better is related to good, running is related to walk)

3. Constructing a model using algorithms to identify the sentiment. The algorithms can be classified as:
   - **Rule-based system** is implemented based on basic scripts and rules. This approach is not very efficient as the script will not be able to support the data with increasing complexity.
   - **Automatic system** uses machine learning techniques to learn from data. Here, a classifier is used to input the text and get the desired output (positive, negative and neutral).
   - **Hybrid system** is a mix of both approaches is right proportion

D. Application

1. **Healthcare**: Sentiment analysis is used to identify positive or negative effects of treatments and drugs on people. It is implemented on textual content found on social media (Facebook, Twitter) and evaluated for results.
2. **Financial Sector**: Stock market trends and emotions of people towards stock is evaluated.
3. **Politics**: Popularity of different leaders and outcome of electoral can be determined using Sentiment Analysis.
4. **Hospitality and Tourism**: Sentiment analysis is used to understand user's perception of services provided at tourist centres and hotels.
5. **Sales**: Understanding the opinion of customers towards new mobile phones or any other gadgets.

E. Challenges

1. Identifying genuine reviews out of duplicate and spam reviews.
2. Identifying negation of a sentence is a challenge. E.g.: R1- "The hotel is not good", R2- "The hotel should provide good service". In the example, R1 indicates explicit negation and R2 indicates implicit negation. Although "good service" is a positive term, the entire sentence is a negative review.
3. Identifying the time when the review was made. E.g.: "The hotel was bad in 2010, but it has got better in 2015".
4. Identifying words that have bi-polar meaning.

References

[1] https://www.researchgate.net/publication/301649777_Analyzing_Scientific_Papers_Based
    _on_Sentiment_Analysis_First_Draft
[2] https://ieeexplore.ieee.org/document/8399738
[3] https://monkeylearn.com/sentiment-analysis/
[4] http://nitin-panwar.github.io/Text-prepration-before-Sentiment-analysis/