ASSIGNMENT-3
SUMMARY

Name: Shrey Amin

Student ID: B00822245

Course: CSCI-5408

Topic: Sentiment Analysis

Sentiment Analysis

Sentiment analysis is widely used concept of natural language processing (NLP), especially as a part of social media analysis for any domain. This technique is used in studying texts, like posts and customer reviews uploaded by users on social networking portals, blogging platforms and websites [3]. This method helps in understanding emotions associated with the opinions users have about a product, service, event, person or idea. So, it is also referred as opining mining. The primary objective of this method is identifying the sentiment of the users and classify whether a given piece of text is positive, negative or neutral.

Sentiment analysis algorithms are used in classifying tweets, movie reviews, news text in articles and user feedbacks for a product. Moreover, as a classification problem, Sentiment Analysis uses the evaluation metrics of Precision, Recall, F-score, and Accuracy to evaluate the results [2][3]. Sentiment analysis may also involve generating graphs, histograms, word clouds and charts to understand several phrases in a text document.

This methodology involves series of steps. Pre-processing is the first step in text and sentiment classification. The given chunk of text is cleaned to get accurate results. Several techniques are used to improve the output of this algorithm. Some of the widely used pre-processing techniques are stemming, part of speech tagging, removing punctuation, converting text into lowercase and removing stop-words [3]. Then the next step is to transform the words such that the transformed word can be used to determine the meaning of the piece of text accurately. After cleaning is completed a feature set is generated and sentiment analysis is performed using various algorithms. The figure below shows the different ways to implement sentiment analysis.

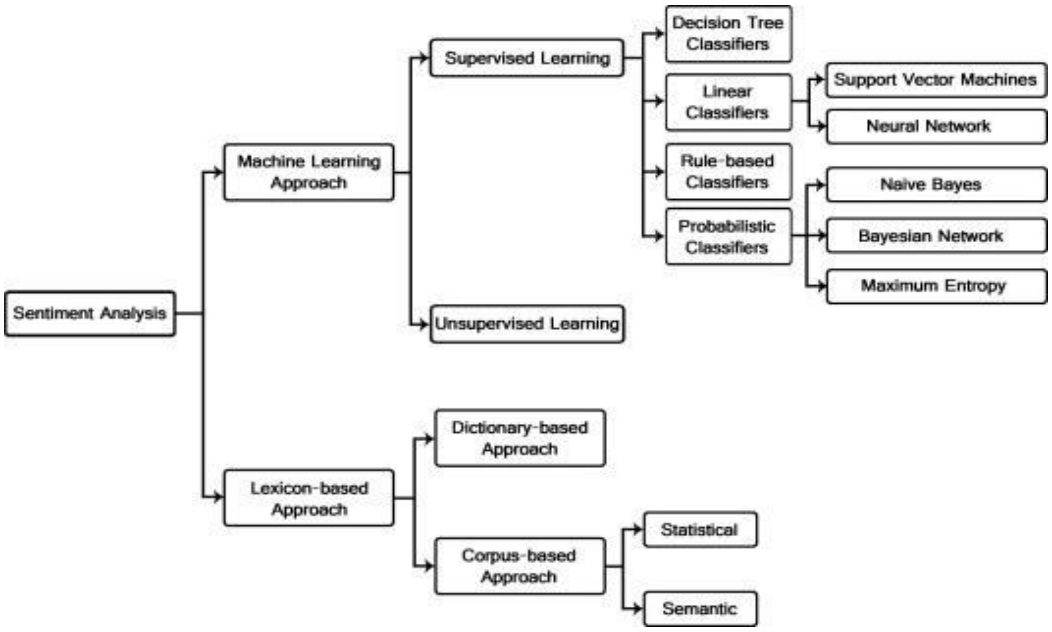

Figure 1 Methods for implementing sentiment analysis [1]

Machine Learning [1]

In this approach supervised and unsupervised learning algorithms are used for determining the sentiments. The supervised methods make use of samples with many features and predefined target labels like positive, negative and neutral. In supervised learning classifiers and algorithms like Support Vector Machine (SVM), Convolutional neural networks (CNN), Naïve Bayes, K-Nearest Neighbours (KNN) and Decision Tree are used. Some of the widely used datasets for training are IMDB movie review dataset and labelled tweets from twitter. Here the dataset collected is divided into training and testing data. A classifier model is created based on the training data. Then the model is tested on the testing data. Further, this trained model can be used to classify or predict sentiment of new text. Apart from this, unsupervised methods are used if the training dataset is not labelled. Here, clustering algorithms are used to group the related words and generate the feature set.

Lexicon Based (NLP) [1]

In this approach lexicons or tokens are generated to analyse the text. It also uses the concept of subjectivity and polarity for the analysis. Polarity means determining a sentiment with a positive or a negative value. Basically, dictionary based, and corpus based are two algorithms used in this method. The dictionary-based approach depends on finding opinion words or generating tokens, and then comparing it with positive and negative words stored in the dictionary. The corpus-based approach is carried out using statistical or semantic methods [1]. Unlike machine learning algorithms lexicon-based approach does not require training data. The disadvantage of this approach is that it difficult to have large number of words that can reflect positive, negative or neutral sentiment.

All in all, machine learning approach is more efficient compare to lexicon-based method as it considers the large labelled training dataset that is used to classifying the text. Though Lexicon based approach uses part of speech (POS) tagging to identify the verb, adjective, noun etc. from the given sentence, still high accuracy is not always achieved. Sometimes this method cannot consider the contextual meaning of a sentence. In such instance a machine learning model tend to perform better. Further, performing sentiment analysis of text with sarcasm is difficult and getting high accuracy is the biggest challenge.

References

[1] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey", *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014. Available: 10.1016/j.asej.2014.04.011.

[2] S. Symeonidis, "5 Things You Need to Know about Sentiment Analysis and Classification", *Kdnuggets.com*, 2019. [Online]. Available: https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html.

[3] M. Kamal Hassan, S. Prasanth Shakthi and R. Sasikala, "Sentimental analysis of Amazon reviews using naïve bayes on laptop products with MongoDB and R", *IOP Conference Series: Materials Science and Engineering*, vol. 263, p. 042090, 2017. Available: 10.1088/1757-899x/263/4/042090.