Sentiment Analysis-Applications and Methods

[1]Ms. Sanjukta Chakraborty, [2]Prof. Samir Kumar Bandyopadhyay

[1]Research Scholar, [2]Professor
[1]Department of Computer Science and Engineering, Seacom Skills University
[2]Department of Computer Science and Engineering, University of Calcutta

***Abstract*: Text data is the main source of communication in several social networking sites. Emotion of a human being is clear only if it is used in combination with gestures and facial expression instead of text. It is important to recognize the sentimental purpose present behind the text if proper analysis is done through the text. Several researches are ongoing to make computer and human interaction as effective as it can detect the sentimental state of a person. This paper reviews exiting work done in different areas of sentiment analysis.**

***Index Terms*: Lexicon, Text Analysis, Machine Learning and Rule-Based System.**

I. Introduction

The root of sentiment analysis is the study of public opinion analysis at the beginning of 20th century and in the text subjectivity analysis performed by the computational linguistics community in 1990's.

Sentiment analysis has now reviewed from analysing online product reviews to social media texts from Twitter and Facebook. It also analysed stock markets, elections, disasters, medicine, software engineering, etc.

The sentiment is analyzed in the entire document as positive, negative or objective.

Sentiment analysis is a series of methods, techniques, and tools about detecting and extracting subjective information, such as opinion and attitudes, from language [1-2]. Sentiment analysis has been done about opinion polarity, i.e., whether someone has positive, neutral, or negative opinion [3]. The objective of sentiment analysis has typically been a product or a service whose review has been made public on the Internet. This might explain why sentiment analysis and opinion mining are often used as synonyms, although it is more accurate to view sentiments as emotionally loaded opinions.

Sentiment analysis is broadly classified into three categories namely, document level, sentence level and aspect based. The goal of document level sentiment classification is determining the overall sentiment of a given review document. Sentence level analysis finds the text at the level of subjective and objective nature. Aspect based approach is more pinpointed as it splits the entire document into various aspects (entities) and sentiment analysis is carried out on each entity to find out the overall polarity.

Sentiment analysis determines the attitude of a speaker or a writer with respect to some topic or the overall understanding of a document. Internet usage and exchange of public opinion is now the driving force behind Sentiment Analysis today. Internet is a huge repository of structured and unstructured data.

Sentiment analysis is broadly classified the applications into the following categories.

a. Applications to Review-Related Websites
b. Applications as a Sub-Component Technology
c. Applications in Business and Government Intelligence
d. Applications across Different Domains

Sentiment Analysis approaches extract positive and negative sentiment bearing words from a text and classify the text as positive, negative or else objective. Objective means if it cannot find any sentiment bearing words.

There are some challenges in sentiment analysis. For example, consider the following sentence:

How can anyone sit to see this movie?

The above sentence does not carry any negative sentiment bearing words although it is a negative sentence. So it is more important to understand semantic of the sentence rather than syntax of the sentence.

Many words change polarity from domain to domain. Consider the following sentence:

Go and read the book.

In the domain of book, it is positive sentence but in case of domain of movie, it is negative sentence since it is assumed that book is not read.

Sometimes context of the sentence has to be analysed. In the following sentence:

The story of film is good but presentation is bad.

In the above sentence, both "good" as well as "bad" words are used but the sentence is negative based on the context of the sentence.

Sometimes user opinion may change the sentiment thoroughly. Consider the following example:

I saw a good film but it is impractical.

The first part of the sentence is positive but second part is negative.

Often it is required to know particular word for detecting sentiments. Consider the following examples:

He is a Frankenstein.

Just finished Doctor Zhivago for the first time.

The first sentence is a negative sentiment whereas the second one indicates a positive sentiment. It is required to know about Frankenstein and Doctor Zhivago to find out the sentiment.

This is to differentiate between opinionated and non-opinionated text. Consider the following examples:

I hate love stories.

I do not like the movie since I hate stories presented in the movie.

The first one is an objective fact whereas the second sentence tells the opinion about a particular movie.

A text or sentence may have multiple entities. Consider the following examples.

Samsung is better than Nokia

Ram defeated Hari in football.

The examples are positive for Samsung and Ram respectively but negative for Nokia and Hari.

It is possible to express opinion negative without the explicit use of any negative word.

A detailed study has been carried out on the various techniques for sentiment analysis. The study reviewed the recent works that has been worked out with various techniques.

II. Literature Survey

Analysis of Sentiment or Opinion mining by researchers is one of the fastest research topic in current scenario of research [1].

Authors discuss the most fundamental problem in sentiment analysis, the sentiment polarity categorization, by considering a dataset containing over 5 million product reviews from Amazon.com with the products belonging to four categories: beauty, books, electronics and home [1].

Authors investigated the predictive power behind the language of food on social media. The dataset contains over three million food-related posts from Twitter and demonstrate that many population characteristics can be directly predicted from this data[2].

Some researchers generated Twitter sentiment indices by analysing a stream of Twitter messages and categorising messages in terms of emoticons, pictorial representations of facial expressions in messages. Based on emoticons they generated daily indices. Then they explored the time-series properties of these indices by focusing on seasonal and cyclical patterns [3].

Authors chosen a particular global ecommerce platform (eBay) and a particular global social media platform (Twitter). They quantified the characteristics of the two individual trends as well as the correlations between it [3].

Authors analysed how Dengue epidemic is reflected on Twitter and to what extent that information can be used for the sake of surveillance. They proposed an active methodology for reducing the disease. It is based on four dimensions: volume, location, time and public perception in tweeter data. The method enables to filter out content that is not relevant for the sake of Dengue surveillance[1].

Some other researchers presented an overview of eight publicly available and manually annotated evaluation datasets for Twitter sentiment analysis. It is difficult to assess most of the datasets for sentiment analysis at target level since it lacks of distinctive sentiment annotations among the tweets and the entities contained in them[3].

Authors analysed several surveys on consumer confidence and political opinion over the 2008 to 2009 period, and find they correlate to sentiment word frequencies in contemporaneous Twitter messages[4].

Researchers introduced the novel approach of exploiting the user influence factor in order to predict the outcome of an election result. Extraction opinions using direct and indirect features of Twitter data based on Support Vector Machines (SVM), Naive Bayes, Maximum Entropy and Artificial Neural Networks based supervised classifiers are used[5].

Authors used a large amount of data collected from Twitter, they presented an in-depth comparison of three measures of influence: in degree, retweets, and mentions. Based on these measures, investigated the dynamics of user influence across topics and time. They made several interesting observations[2].

Authors investigated whether removing stop words helps or hampers the effectiveness of Twitter sentiment classification methods. They observe how removing stop words affects two well-known supervised sentiment classification methods [1].

This paper employs a new dataset of over 500GB of politics-related Tweets from the final months of the 2012 presidential campaign to interpolate and predict state-level polling at the daily level. By modelling the correlations between existing state-level polls and the textual content of state-located Twitter data using a new combination of time-series cross-sectional methods plus Bayesian shrinkage and model averaging, it is shown through forward-in-time out-of-sample testing that the textual content of Twitter data can predict changes in fully representative opinion polls with a precision currently unfeasible with existing polling data[3].

Authors compared seven opinion lexicons on six sentiment datasets (movie reviews and tweets) conducted. Results suggested that increasing the lexicon size by semantic expansion as well as assigning an interval value to the words of the opinion lexicon significantly increases the classification performance on short texts (e.g. tweets)[4].

Researchers proposed a sliding window Kappa statistic for evaluation in time-changing data streams. Using that statistic they performed a study on Twitter data using learning algorithms for data streams[5].

Authors analysed that the daily variation of election voting intentions expressed on Twitter to evaluate the effect of different campaign messages, measured through the hand coding of parties and leader's official Twitter account[4].

Authors examined the use of information embedded in the Twitter stream to track rapidly-evolving public sentiment with respect to H1N1 or swine flu, and track and measure actual disease activity. They also showed that Twitter can be used as a measure of public interest or concern about health-related events. Their results showed that estimates of influenza-like illness derived from Twitter chatter accurately track reported disease levels[1].

Authors attempted to understand whether Twitter data mining can predict about climate change to eliminate related hazards[4].

Researchers detected whether user comments on an on-line newspapers reflect external indicators of public acceptance (e.g. vote intention). The paper outlines the approach used to identify and classify sentiment in news comments written in Portuguese language and to correlate it to external indicators, and discusses the main results for this case study[4].

Authors proposed study the problem of unsupervised sentiment analysis for categorizing two main emotional signals, i.e., emotion indication and emotion correlation.

Authors introduced model to incorporate prior knowledge for tracking illnesses over times (syndromic surveillance), measuring behavioral risk factors, localizing illnesses by geographic region, and analysing symptoms and medication usage[5].

Some researchers applied LDA-G to several large graphs (with thousands of nodes) from PubMed (a scientific publication repository). They compared LDA-G's quantitative performance on link prediction with two existing approaches: one Bayesian (namely, Infinite Relational Model) and one non-Bayesian (namely, Cross-associations). They demonstrated that LDA-G can discover useful qualitative information[4].

Authors tried to characterize the opinion-mining landscape by proposing a faceted taxonomy of the different aspects of opinion mining. They then surveyed literature and place these in appropriate places in the proposed model. They also proposed a general purpose work flow required from any opinion mining engine[3].

Authors focused on different features with different scopes on face book can be found by multi-level opinion-consistent hidden community(OCHC) framework that proposed in the paper. Communities of opinion-consistent users are clustered Multi-level OCHC model [3].

Authors employed the concept of public opinion field, on which event information and public opinion in text corpus are distinguished. Based on this view, they focused on how does the public opinion affect the evolution of events, proposed a method to measure the influence, and represent it on the event evolution graph[2].

Authors illustrated of a novel opinion mining method underpinned by context-sensitive text mining and inferential language modelling to improve the effectiveness of opinion mining. Their initial experiments showed that the proposed the inferential opinion mining method outperforms the purely lexicon-based opinion finding method in terms of several benchmark measures[5].

Researchers approached OFESP (Opinion Feature Extraction based on Sentiment Patterns) which takes into account the structure characteristics of reviews for higher values of precision and recall. With a self-constructed database of sentiment patterns, OFESP matches each review sentence to obtain its features, and then filters redundant features regarding relevance of the domain, statistics and semantic similarity[4].

Authors compared to the traditional unsupervised alignment model, the proposed model obtains better precision because of the usage of partial supervision.

Researchers proposed a new approach based on opinion mining. Through the reviews customers have posted they can mining the evaluation of various Web site indexes quantitatively. To improve the accuracy of the mining results they used a approach called MRA (mutual reinforcement approach)[2].

Authors proposed method of opinion mining from existing ones. The methods enable credibility evaluation and result conversion using influence of each opinion holder on the Internet and their personal information[3].

Authors explored an idea of extracting real time dataset through provided Graphical User Interface (GUI). The summarization unit would generate opinion mining result in visualized form for further decision making process[4].

Researchers proposed a Leader Rank algorithm to identify opinion leaders based on community discovery and emotion mining methods. The performance of this algorithm is evaluated using real-world datasets and their experiments showed that the identification of interest groups and the emotion property shown in post/reply articles helps to find opinion leaders on Bulletin Board System (BBS)[1].

Authors proposed a opinion mining system which mines useful opinion information from camera reviews by utilizing Semantic Role Labeling (SRL) and polarity computing method. Feature lexicon and sentiment lexicon are constructed to mine features and emotional items[5].

Authors introduced the use of a fuzzy lexicon and fuzzy sets in deciding the degree of positive and negative. Their experimental result showed that the approach is able to extract opinions and present the opinions in a more efficient way[2].

Researchers proposed a framework of feature based opinion mining by using scores which essentially relies on the usage of two main lexiconizing levels, features and polar words[2].

Authors presented a recommendation technique based on opinion mining to propose top ranked books on different discipline of the computer science. They analysed the features on the basis of several characteristics to categorize and reviews of the users[3].

Authors proposed a new system architecture that can automatically analyse the sentiments of these messages. They combined this system with manually annotated data from Twitter, one of the most popular micro blogging platforms, for the task of sentiment analysis[5].

Authors presented a product ranking system using opinion mining techniques. In that system, they considered three issues while calculating product scores: 1) product reviews, 2) product popularity, and 3) product release month[5].

Authors showed Logic Programming, particularly Answer Set Programming (ASP), that can be used to elegantly and efficiently implement the key components of syntax based aspect extraction[4].

Authors described in this research provides topological clustering of the opinion issued from the tweets, each cluster being associated to a prototype and a weight vector, reflecting the relevance of the data belonging to each cluster[1].

Researchers presented a generic and domain independent opinion relevance model for a Social Network user. The Social Opinion Relevance Model (SORM) is able to estimate an opinion's relevance based on twelve different parameters. Compared to other models, SORM's main distinction is its ability to provide customized results according to whom the opinion relevance is being estimated[2].

Authors described functions of a system designed for the behaviour analysis of e-commerce clients. It enables user identification and client behaviour extraction for interacting with web site customers. Their system carries out an evaluation and rating of opinions, and their approach is based on linguistic and the statistic treatment of natural language[4].

Authors discussed problem of different sentiment of the same sentence in different environment. To solve this problem, rule-based classification is used as their machine learning model[3].

Authors analysed data collected from Twitter and investigate the variance that arises from using an automated sentiment analysis tool versus human classification. Their interest particularly, lies in understanding how users' motivation to post messages affects the quality of classification. The data set utilizes Tweets originating from two of the world's leading oil companies, BP America and Saudi Aramco, and other users that follow and mention them, representing the West and Middle East respectively [4].

Researchers discussed that most of the existing methods are processing the reviews in terms of positive and negative comments. But this approach is not enough for a customer to make decision about product. The proposed approach not only finding the positive and negative comments for any product or product features, but also rating them in the order of positivity and negativity. Also, the proposed approach gives the degree of comparisons for a particular product and product features[3].

Authors discussed that in order to get useful data it becomes necessary to apply NLP techniques which make it easy for the people to make decisions at the time of buying products or contracting services[2,6].

III. Different Applications

Sentiment analysis can be used in many areas for various purposes. The most general use of sentiment analysis is in ecommerce activities. Users can submit their experience about shopping and product qualities in websites. Users can also give their ratings or scores of different products. Customers can easily view opinions and recommendation information on whole product as well as specific product features. It is possible to present graphical summary of the overall product and its features to users. Some product survey companies contain 75 millions opinions and reviews worldwide. Sentiment analysis of these helps such websites by converting dissatisfied customers into promoters by analysing this huge volume of opinions. Voice of the Market determines customers feeling about products or services of competitors. It helps in gaining competitive advantage and new product development. It is important to detect this information as early as possible helps in direct and target key marketing campaigns. Real time Sentiment Analysis of customer helps to design new marketing strategies, improve product features and can predict chances of product failure.

Voice of the Customer analyses the reviews and feedback of the customers about product(s). It is a key element of Customer and helps in identifying new opportunities for product inventions. It helps identify functional requirements of the products and some non-functional requirements like performance and cost.

Reputation of a company in market is done by Brand Reputation Management . It is a process and company focused on it properly It creates opportunities for organizations to manage and strengthen brand reputation. It is determined by advertising, public relations and corporate messaging. Sentiment analysis helps in determining company's brand, product or service to the customer online. The performance of Government also can be assessed by Sentiment analysis based on opinions from public.

Sentiment analysis has practical applications in teaching contexts, especially in online courses where so much of the learning and discussing happens as typed text. One possibility is to monitor in real time the online conversations students have in course discussion forums, discussion groups, or social media channels and to analyse the sentiment evident in this text. In face-to-face discussions, instructors can actively and consistently monitor every conversation in terms of students' motivation, mood, and understanding of material. However, instructors in online course conversations are often "late to the party" because many interactions and postings unfold between instructor logins. Using sentiment analysis, instructors could be notified via the messaging service of their choice when sentiment suddenly changes in a course. Quick changes in the mood of a course or conversation can be an important moment in a course, and early notification and intervention can be key to instructional fidelity in such instances. Sentiment Analysis is made on course-related online communications and organizes data on a student-by-student basis can also be done for analysis of the performance of the student(s).

Sentiment analysis offers a potential tool in studying the many communities of practice that exist in educational settings, including professional development communities, informal teacher communities, professional communities, and academic communities. These analyses could inform the study of how communities change and evolve over time, how they differ from site to site (e.g., between teachers from two different districts), or how communities respond to particular events (e.g., introduction of new standards). It is also possible to use sentiment analysis to study individuals and to establish how a single teacher's experience might change as they enter a new community, as they become an established member of that community, and yet again as they become a senior member within the same community.

IV. Different Methods

Research works on Sentiment Analysis can be classified as technique used, view of the text, level of detail of text analysis, rating level, etc. From a technical point of view, approaches are machine learning, lexicon-based, statistical and rule-based approaches. Several learning algorithms are used in machine learning method to determine the sentiment by training on a known dataset. The lexicon-based approach finds sentiment polarity for a review using the semantic orientation of words or sentences in the review. The measure of subjectivity and opinion in text is done by sentiment orientation. The rule-based approach looks for opinion words in a text and then classifies it based on the number of positive and negative words. Statistical models try to cluster head terms into aspects and sentiments into ratings.

Another classification is oriented more on the structure of the text: document level, sentence level or word/feature level classification. For finding sentiment polarity for the whole review aims document level classification in which sentence level or word-level classification can express a sentiment polarity for each sentence of a review and even for each word.

V. Conclusions

In this review works our main focus on applications and some of methodologies for Sentiment Analysis. Several Software are now available for proper analysis but still it has many applications where still no works have been done.

REFERENCES

[1] Basha, S. M., Zhenning, Y., Rajput, D. S., Iyengar, N., & Caytiles, D. R. (2017). Weighted Fuzzy Rule Based Sentiment Prediction Analysis on Tweets. International Journal of Grid and Distributed Computing, 10 (6), 41-54.

[2] Law, D., Gruss, R., & Abrahams, A. S. (2017). Automated defect discovery for dishwasher appliances from online consumer reviews. Expert Systems with Applications, 67, 84-94.

[3] Nguyen, D. Q., Nguyen, D. Q., Vu, T., & Pham, S. B. (2014). Sentiment classification on polarity reviews: an empirical study using rating-based features. In Proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis (pp. 128-135).

[4] Tiwari, P., Mishra, B. K., Kumar, S., & Kumar, V. (2017). Implementation of n-gram methodology for rotten tomatoes review dataset sentiment analysis. International Journal of Knowledge Discovery in Bioinformatics (IJKDB), 7 (1), 30-41.

[5] Arun, K., Srinagesh, A., & Ramesh, M. (2017). Twitter Sentiment Analysis on Demonetization tweets in India Using R language. International Journal of Computer Engineering in Research Trends, 4 (6), 252-258.

[6] Anta, A. F., Chiroque, L. N., Morere, P., & Santos, A. (2013). Sentiment analysis and topic detection of Spanish tweets: A comparative study of of NLP techniques. Procesamiento del lenguaje natural, 50, 45-52.