Sentiment Analysis : A Survey

U. Aggarwal[1*], G. Aggarwal[2]

[1]Department of Computer Science, Jagannath University, NCR, Haryana, India
[2]Department of Computer Science, Jagannath University, NCR, Haryana, India

Corresponding Author: uma.aggarwal1992@gmail.com

Abstract— Sentiment analysis is a technique to analyze people's opinion on given topics such as political, social, and economical or review on product. An enormous amount of information and opinion is available online which is helpful for researchers to analyze in fields such as market research political issue, business intelligence, online shopping etc. This paper presents a survey which covers Sentiment Analysis, literature survey, classification, application and challenges.

Keywords— Sentiment Analysis , Classification

I. INTRODUCTION

Sentimental Analysis or opinion mining is a technique used to study people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions expressed in an article, issues, events, and topics. It focuses on identifying whether a given text is subjective or objective and if it is subjective, then whether it is negative or positive. Sentiments can be expressed on almost anything and everything, e.g., a product, a service, an event, or an individual view. It is one of the most widely studied applications of Natural Language Processing (NLP), Data Mining[9] and Machine Learning. NLP is a computerized approach used for analyzing naturally occurring data. With the rapid growth of Internet, people have different platforms such as twitter, facebook, Google+ etc to express their views, emotions and sentiments towards products, people and life. So, to give reviews and get feedback of different topics social networking sites are the best resource for tracking and analyzing public sentiment. Thus, the Internet is a vast resource of opinion rich textual data. Hence, it seems to be huge repository of structured and unstructured data. The aim of Sentiment Analysis is to withstrain this data in order to obtain critical information regarding public opinion, an emotions that help to make smarter business decisions, political campaigns and better product consumption.

The rest of paper is organized as follows. Section 2 provides about literature survey on sentiment analysis.Section 3 gives an idea of levels of sentiment analysis. Section 4 explains about applications of sentiment analysis. Section 5 gives brief knowledge of challenges in sentiment analysis. Section 6 will be conclusion. Section 7 will be references.

II. LITERATURE REVIEW

It has been widely researched topic. The term 'social network' was first coined by John Arundel Barnes. In 1954, he presented result of nearly two year of studies on the composition of classes and social groups in the town of Bremnes[1]. In year 2012 , G. Vinodhini has explained the technique used for sentiment classification which includes Naïve Bayes, the basic idea is to estimate the probability of categories given a document by using the joint probability of words and categories, and Statistical classification method based on the structural risk minimization principle from the computational learning theory (SVM), Centroid Classification, Winnow, K-nearest neighbour Method are well-known as online mistaken-driven method, and Ensemble technique. These techniques combine several base classification output to generate final output [2].

Dhanashri Chafale and Amit Pimpalkar, explained that instead of stating the complete document into positive or negative, sentiments for a particular topic from a document can be expressed as positive or negative. Moreover, they defined the basic issue in sentiment analysis which is getting the sentiment expressed in texts whether the sentiment shows positive or negative opinion [3].

Taylor et al. shows that designing a system of opining mining for tourism will help a lot in tourism industry [4]. They also proposed a system to solve problem in Lake Distract tourism industry. The authors proposed a method using naive bayes and modified k means clustering on mobile review. According to this paper, this proposed method is found better than naive bayes and support vector machine techniques individually [5].

In Moraes, Valiati and Neto focused on achieved good classification accuracies by comparing between SVM and ANN. Also, their experiments examined all methods as a function in bag-of-words (uni grams) approach in particular terms [6].

III.   LEVEL OF SENTIMENT ANALYSIS

A.   Document Level Sentiment Analysis

Document level sentiment classification focus on classify the entire document as positive or negative. The main issue in the document level classification is not all the sentence in a document have relevance in expressing the opinion about an entity. Hence, irrelevant sentences should be eliminated from the processing. Supervised or unsupervised learning methods are used in the document level classification. Supervised learning algorithm such as Support Vector Machine, naive Bayesian are used to train the system. The unsupervised learning is done by extracting opinion of words inside a document. The document level sentiment classification has both advantages and disadvantages. Advantage is we can get an overall polarity of opinion text about a particular entity from a document. Whereas, Disadvantage is different emotions about different features of an entity could not be extracted separately.

B.   Sentence Level Sentiment Analysis

Sentence level sentiment analysis is used to get the polarity of each sentence is calculated. The similar document level classification methods are used for sentence level classification problem. It helps to find Objective and subjective sentence. In subjective sentences opinion words help in determining the sentiment of the entity. After that polarity classification is done on positive and negative classes.

C.   Phase Level Sentiment Analysis

The phrase level sentiment classification is a more specific approach for opinion mining. The phrases that contain opinion words are searched and classification is done on that. This can be advantageous or disadvantageous. In some case, the exact opinion of an entity can be correctly extracted.
 in other cases, where contextual polarity matters, the result may not be accurate.

IV.   APPLICATIONS OF SENTIMENT ANALYSIS

It should include important findings discussed briefly. Wherever necessary, elaborate on the tables and figures without repeating their contents. Interpret the findings in view of the results obtained in this and in past studies on this topic. State the conclusions in a few sentences at the end of the paper. However, valid colored photographs can also be published.

A.   Summarization of Reviews

Internet is hub of reviews and feedbacks of almost every topic which includes product reviews, feedbacks on political or social etc. Thus a sentiment engine is required that can extract sentiments about a particular entity. It will provide an overall feedback or rating for the given topic. A consumer may not be able to decide about product through reading all reviews and make an informed decision about product and the manufacturer may not be able to keep track of consumer opinion.

B.   Applications in Business Intelligence

Nowadays people tend to look upon reviews of products online before buying them. Many businesses take decision on basis of the online opinion of their product. Thus, Sentiment Analysis plays a vital role in businesses. Businesses require online reviews in order to improve their products.
Sentiment Analysis is also be used for trend prediction. On basis of public opinion, important data regarding sales, trends and customer satisfaction can be extracted.

C.   Recommender System

Sentiment Analysis helps in knowing individual's review on a product or issue. On based of that customers' take decision and further predict what impact that topic has on other domains. For example, if we analysis public opinion on demonetization in India, we can predict peoples opinion on this. Moreover we can predict what effect demonetization has on social and economic fronts. This paper explains how recommender system helps to provide best product to customer [10].

D.   Managing brand reputation

Brand Reputation is managing by knowing your reputation in market. Opinions of customers effect one's reputation. So to maintain ones reputations company asks customers reviews on basis if which they improve their product. This gives opportunities to organizations to manage and strengthen brand reputation.

E.   Opinion analysis Politics

Sentiment analysis helps in analyzing opinion on issues and comments of people in political blogs. To improve the level of the information available to voters, the position of public figures, through which support or oppose, can also be determined. Mullen and Malouf explain a statistical sentiment analysis method on political discussion group postings used to judge whether there is opposing political viewpoint to the original post [8]. Researchers have pointed out some research challenges such as identifying of opinion holder, associated opinion with issues, identifying public figures and legislation.

F. Government Intelligence

Government intelligence is one of the applications of sentiment analysis. It helps efficient rule making by automatic analyzing the opinions of people about pending policies or government-regulation proposals. Moreover it can be used for tracking the citizen's opinion about a new scheme, predicting the exit polls or success of a new legislative reform to be introduced and determine the mood of the public towards a scandal or controversy.

V. CHALLENGES OF SENTIMENT ANALYSIS

Sentiment analysis is used to classify text as positive, negative. So it can be called as text classification task. So there can be many factors that make sentiment analysis difficult task as different people using different language or emotion which makes analysis task difficult. The following factors are discussed in brief:

A. Pronoun or noun resolution

The problem of identifying what a pronoun, or a noun phrase refers to coreference resolution. The object used for opinion is expressed in first sentence or early part of sentence. The other part of sentence may also contain that object referenced as pronoun like it, this, that, he, she etc.
For example, "We watched the match and went to lunch; it was awful." What does "It" refer to? Coreference resolution may be useful for the topic aspect based sentiment analysis. To resolve this pronoun is a complex task.

B. Time of Review

The time of reviews is important for sentiment analysis. The reviewer point of view may change with time as window vista was considered best but now people are switching to windows 7. So, this change of opinions with time improves the performance of the sentiment analysis system. Moreover, it helps us to examine if rating of any product or people opinion on any topic is changed or not.

C. Sarcasm Handling

Negation has a important role in changing the polarity of the associated adjective and change polarity of text. Text may contain Sarcastic and ironic sentences which has negative sense. For example, "What a wonderful bike, it stopped working second day." In this case, positive words have negative meaning. So, such cases are hard to identify which can lead to erroneous opinion mining.

D. World Knowledge Required

Knowledge about worlds' facts, events, and people required may to correctly classify the text. The text contains another entity to refer to one entity. For example, "She is as good as Snow White" The system without world knowledge classifies above sentence as positive due to the word "good", but it is an objective sentence because one has to know about 'Snow White'.

E. Domain Generalization

Some words exhibit different polarities when used in different domains. The accuracy of sentiment classification is depicted by the domain of the items in which it is applied. As, there are many words whose meaning is different in different domains. For example, "Go do your homework." This sentence has positive sentiment in homework domain while it shows negative sentiment for movie domain.

F. Expectations

Some text contains sentences starts with different context while end at different context. For
Example, "The cast was good, actors performed great, but I don't like it." In this review the last sentence makes the whole review negative. If term frequency is considered this statements would be classified as positive due to more positive words in review. But, it's difficult to determine meaning of overall review. This is another challenging task in sentiment analysis.

G. Negation handling

The small differences between two pieces of text changes complete meaning of text. In Sentiment analysis, Negation handling is difficult tasks as polarity gets reversed. "The city is beautiful" is different from "the city is not beautiful". In above example "beautiful" is a positive word but "not" is a negative word. Thus handling sentences such as "not beautiful" makes sentiment analysis a difficult task .Negation is also expressed in sarcasm or implicit sentences which doesn't contain any negative words.

VI. CONCLUSION

Sentiment Analysis is one of the important research areas as it summarizes opinions and reviews of public. This survey highlights the main idea behind Sentiment Analysis and explains literature review, Sentiment Classification, challenges in detail. Sentiment analysis is one of the active reasech areas and several interesting works have been done in this field. Still, a fully useful and highly efficient system has not been introduced till now. But business organizations and academics are working hard to find the best system for sentiment analysis. Sentiment analysis helps in decision making and knowing people review by analyzing or giving rating to their views such as product reviews .

VII. REFERENCES

[1] J. A. Barnes, "*Class and Committees in a Norwegian Island Parish, Human Relations*", Human Relations, Vol.7, Issue.1, pp. 39-58, 1954.

[2]   G. Vinodhini , RM. Chandrashekharan, "*Sentiment Analysis and Opinion Mining : A Survey*" , International Journal of Advanced Research in Computer Science and Software Engineering, Vol.2, Issue. 6, pp.281-292, 2012.

[3]   Dhanashri Chafale and Amit Pimpalkar, "*Review on Developing Corpora for Sentiment Analysis Using Plutchik's Wheel of Emotions with Fuzzy Logic*", International Journal of Computer Sciences and Engineering, Vol.2, Issue.10, pp.14-18, 2014.

[4]   E. Marrese-Taylor, J. D. Velasquez, F. Bravo-Marquez, *"Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web"*, In the Proceedings of the 2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI) and Intelligent Agent Technology (IAT), CA, pp. 261–264, 2013.

[5]   Preety , Sunny Dahiya , *"Sentement Analysis using svm and naive bayes algorithm"*, International Journal of Computer Science and Mobile Computing, Vol.4, Issue.9, pp.212-219,2015.

[6]   R. Moraes, J. F. Valiati, W. P. Gaviao Neto, *"Document-level sentiment classification: An empirical comparison between SVM and ANN"*, Expert System with Application, Vol.40, Issue.2, pp.621–633, 2013.

[7]   Mullen T. and Malouf R., *"Taking sides: User classification for informal online political discourse"*, Internet Research, Vol.18, Issues. 2, pp.177–190, 2008.

[8]   AR. PonPeriasamy, G. Vijayasree, "*Data Mining Techniques for Customer Relationship Managemen*t", International Journal Of Computer Sciences and Engineering, Vol.5, Issue.4, pp.120-126 , 2017.

[9]   R. Keshav, Arvind Swaminathan S, Goutham R, Naren J, *"Content based Recommender System on Customer Reviews using Sentiment Classification Algorithms"*, International Journal of Computer Science and Information Technologies, Vol. 5,Issue.3 ,pp.4782-4787, 2014.

Authors Profile

Uma Aggarwal pursed Bachelor of technology from banasthali University, Banasthali in 2015 and is currently pursuing Master of technology from Jagannath University.

Dr. Gaurav Aggarwal received M.Tech. (Computer Science and Engineering) degrees from Maharshi Dayanand University in 2008. Presently, he is working as an Assistant Professor in Computer Science and Engineering Department in Jagannath University, NCR, Haryana, India. Her areas of interest are software Reliability and Neural Networks.