International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 10, October-2018

EXPLORATION AND EFFECTIVENESS OF SENTIMENT ANALYSIS: A REVIEW

Priyanka sharma[1] Arvind kalia[2] Naveen kumar[3]

[1]*Department of Computer Science Himachal Pradesh University, Shimla India,*
[2]*Department of Computer Science Himachal Pradesh University, Shimla India,*
[3]*Department of Computer Science Himachal Pradesh University, Shimla India,*

Abstract— Data analytics is becoming a golden step in assessing the real opinions and emotions of people. Sentiment analysis is a type of data analytics which inspects the social media data to determine the opinions of people in an organized way. Sentiment analysis has come a long way since determining the opinions merely from adjectives to the machine learning approach. It is the most effective way to inspect the sentiments which makes it one of important process to get insight into intentional sphere of the customers to gain an edge in the competitive world.

Keywords— Sentiment analysis, Opinion mining, Twitter, Natural language processing, effectiveness of sentiment analysis.

I. INTRODUCTION

Sentiment analysis is a field of study that deals with the analysis of people's opinions, feelings, sentiments, emotions from the text, which people express towards a product, services, individuals, events etc. Sentiment analysis is also known by other names, such as opinion mining, sentiment mining, subjectivity analysis. However, they all follow the same semantics to determine the sentiments [1]. The term sentiment analysis firstly appeared in the work of Nasukawa and Yi [2] they defined the term sentiment analysis as a task to find sentiment expressions of a subject and inspecting the polarity of these sentiments. Besides adjectives other form of words like nouns, verbs, adverbs can also be used to express the sentiments. However, research in this area results back in 1997 and 2001. Hatzivassiloglou et.al. [3] identified and validated the positive or negative semantic orientation of the conjoined adjectives. They used a clustering algorithm to separate the adjectives and then labelled them with positive or negative. Since the focus of the work was adjectives, sentiments can be written using many words other than the adjectives. Work presented in these years mostly recognized text according to its subjectivity and then determined its orientation.

Moringa et.al [4] in 2002 presented a new framework for mining the reputations of the product with internet as a source for people's opinion. They collected these opinions from the web pages and used text mining techniques to obtain reputation of the products. They performed four kind of tasks on the data collected: to identify the characteristics words they people choose to express their sentiments, point out the co-occurrence of the words, identify typical sentences and perform the analysis on target categories. Since 2002 content on the internet has increased many folds. Opiniated text is now available in many forms, over the entire social platforms. Micro blogging websites like Twitter is the hub for expressing any kind of sentiments and opinions. Research work in sentiment analysis area has also speeded up. It is one of the most active area in research. Research problems in this area is investigated at many levels.

Pang et.al [5] classified the entire document into positive or negative sentiments. They examined whether to classify document simply based on topics or whether a special sentiment -categorization should be developed. Document level analysis assumes that whole document provides sentiment about a specific document only. Sentence level analysis deals with sentences. Whether each sentence specify a sentiment or not. It is also known as subjectivity classification. Wiebe et.al [6] distinguished sentences, those were used to present factual information. Sentences were categorized info subjective and objective sentences. Objective sentences express factual information and subjective sentences expresses real opinions. But all subjective sentences cannot express all the sentiments, objective sentence also imply opinions. Aspect level analysis is a type at finer level. This kind of analysis is based on features. Instead of dealing with construct of the language aspect level analysis directly act upon opinions. Hu and Liu [7] identified user reviews about a product on the basis of the feature mentioned about that product, e.g. picture quality of a camera. With such feature-based distinction another customer can easily identifies what other person feels about the same feature. However, real time data analysis also requires the monitoring of the customer reviews to act upon it whenever it is present.

Sentiments and opinions are provided about an entity. These are the user reviews that he/she express. It can be in the form of positive or negative reviews or in numerical values in the form of scores. These sentiments can be classified using several techniques. Medhat et.al [8] presented all techniques and methods under them to classify the sentiments.

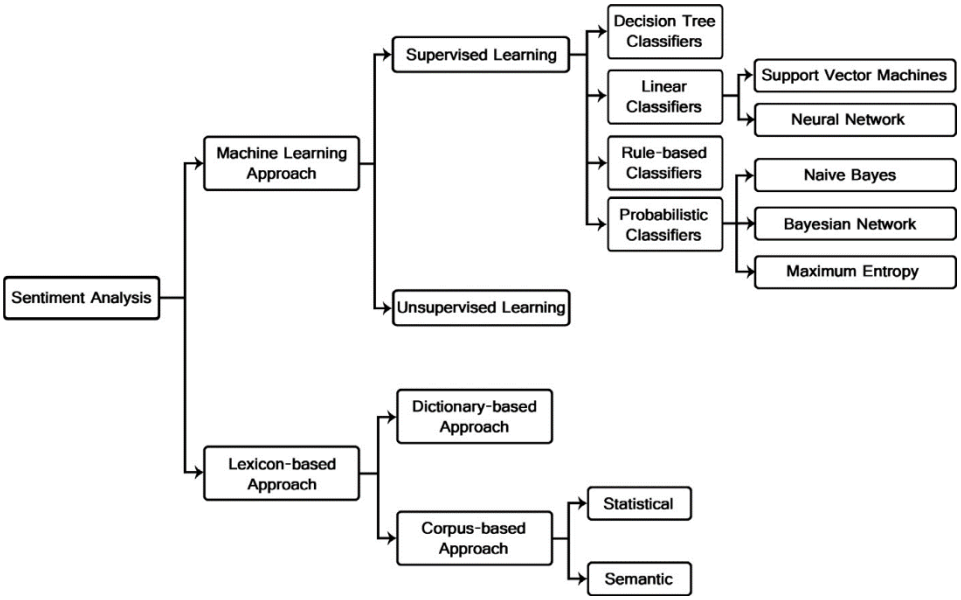

Figure 1 Sentiment classification techniques [8].

Supervised learning approach uses methods like naïve Bayes, support vector machines. In supervised learning features like frequency, parts of speech are used. User reviews expressing sentiments have specific words which are identified as opinions. A list of such words can be created. Positive and negative words collectively are called sentiment lexicon. Sentiment analysis performed using lexicon is called dictionary-based approach. Miller et.al. [9] used WordNet an online lexical reference which was based on psychological theories. Wordnet divides the lexicon into four categories nouns, verbs, modifiers and function words.

Opinions are centre of all human activities and they influence our behaviour. Organizations and businesses always want to know what consumer thinks about their services thus analysis of such opinions have real time applications. Liu et.al. [10] proposed a model to predict sales performance. Miller et.al. [11] explained that sentiments of a blog get affected not only by immediate parent but also its position in the cascade. Analysis results can be used by organizations in the improvement of their businesses to provide effective services to the customer to gain an edge in the competitive market.

This review paper has been divided into three sections. First section explains the detail review study of the sentiment analysis. Second section discuss the effectiveness of the sentiment analysis approach used in various topics. Third section concludes the review paper.

II. EFFECTIVENESS OF SENTIMENT ANALYSIS

Chamlertwat W et.al. [12] reported an exploratory study using sentiment analysis performed on micro-blog. Study shows that sentiment analysis on micro-blog like Twitter provides encouraging results. They studied the reviews of different smartphones. Top three brands which received highest confidence scores were iPhone, Nexus and Blackberry, while Samsung and HTC received good reviews in certain features like camera. They validated the results by asking experts in this area on the basis of scores from 5to 1. Results showed that experts strongly complied with the results obtained from analysis of reviews from Twitter.

Brendan O'Connor et.al. [13] measured the public opinion from polls with sentiment analysis of the text from the popular site Twitter. They measured the link between the two through time and polling data. They analysed several surveys on consumer confidence and political opinion over the years 2008 to 2009 period. Over the several datasets correlation was as high as 80%. Analysis captured large-scale trends. Their results show that tiresome polls can be supplemented with sentiment analysis. If more advance NLP techniques are used, results can be better.

Andrea Ceron et.al. [14] applied a proposed approach to two different scenarios. They analysed online popularity of Italian political leaders throughout 2011, in second scenario they analysed voting intention of French internet users in presidential poll. Their analysis showed astounding correlation between social media and the results of traditional mass surveys. They pursued dynamic approach for French electoral campaign, collected 244,000 tweets. In polls Hollande won the ballot against Sarkozy by a margin of 51.64% of votes. According to online analysis of tweets related to it also predicted the win for Hollande with 54.9% votes. This phenomenon was also true for other scenarios too. This analysis shows that social media can provide accurate predictions and results.

Yang Yu et.al. [15] performed a study to investigate the impact of social media over conventional media on firm stock performances using sentiment analysis. They applied advance sentiment analysis technique to analyse the overall sentiments of each media from over no. of resources. Their findings show that social media has more impact on firm stock performances. With a sample of 52,746 messages from more than 800 firms from various social media platforms. Positive blog posts have more prominent impact on the return of the stocks, negative posts have more negative impact on the returns. Sentiments of Twitter posts can effectively point out the risk related to specific stocks. This shows that text analysis can be a better way to gather important information, which can contribute towards better understanding of stock returns and risks associated with them.

Alvaro Ortigosa Et.al. [16] demonstrated that it is possible to extract information from the student's sentiments from the messages they post on the Facebook. They presented new method for sentiment analysis, which get information about polarity of user's sentiments. Regular record of these sentiments over the time can detect the emotional changes in student's behaviour. They developed SentBuk, a Facebook application that collects messages, comments, likes on the user profile and maintains a sentiment profile of that user. Which can be communicated with user using an interface. This has relatively huge application value. In can, be used in the context of e-learning. SentBuk analyses messages of the students, weekly, compares it with students' regular pattern to detect the changes. This can detect the positive, negative sentiments of a student regarding a course or an activity.

Felix Greaves, Daniel Ramirez-Cano et.al. [17] attempted to use machine learning to understand patients' comments about their care, they used sentiment analysis to categorize the reviews of patients into positive, negative according to their feedback about health care. They performed analysis on 6412 online comments on health service website. They compared the results obtained from analysis with the national inpatient survey results. They found significant associations between analysis and the survey. The relationship between two analysis was reflected as Kappa statistics between .40 and .74. This shows that sentiment analysis of social data has significant association with real sentiments of the people.

Flora Poecze et.al. [18] presents the results of an analysis on self-marketing techniques of YouTube gamers. They analysed their Social media content like on Facebook with their strategies used form subscription on YouTube. To perform sentiment analysis on Facebook content of these gamers supervised learning method "k nearest neighbour" was used. Accuracy of analysis close to 82.3% was achieved. Results of the study show that solely relying on the YouTube metrics, instead sentiment analysis of social media posts should be used for increasing the customer growth.

Paramita Ray, Amlan Chakrabarti et.al. [19] analysed the after impact of "demonetization" a sudden step of Government of India. Sentiment analysis was performed using Twitter data, user's reviews were inspected for after affect. Tweets was collected in two phases: firstly, after November 8, 2016 and in second phase on December 14, 2016. Score for eight cities was maintained and compared. ANOVA statistical method was employed to determine the difference between multiple sets of data. Effectiveness of sentiment analysis was measured in two ways: classification performance was measured in terms of precision(P), recall (R) and F- on positive reviews. Accuracy shown by the method was 70%. In second method co-relation coefficient matrix for eight cities was calculated. Results indicated that primary concerns of different groups about the adverse impact of demonetization was not confirm by the user reviews.

III. CONCLUSIONS

Sentiment analysis is the fastest and most effective process to gather insights into opinions, emotions, thinking of people through the reviews, feedbacks that they share over the social media. With the use of machine learning techniques, it has become most successful and productive process in the field of data analytics. Analysis of the user reviews can prove strategic in different areas which connects users to the service providers. Effectiveness of the sentiment analysis can be used to improve the output and growth of the organizations in the business and marketing world.

REFERENCES

[1] Liu Bing, "Sentiment Analysis and Opinion Mining", Toronto Canada, Morgan &Claypool Publishers, April 2012, pp 20-100 [Online]. Available at: Morgan Claypool, https://www.morganclaypool.com
[2] Nasukawa Tetsuya, Yi Jeonghee, "Sentiment Analysis: capturing favorability using natural language processing", 2003 pp 70-77 [Online]. Available at: https://dl.acm.org
[3] Hatzivassiloglou Vasileios, McKeown R Kathleen, "Predicting the semantic orientation of adjectives", 1997 pp 174-181 [Online]. Available at: https://dl.acm.org
[4] Moringa Satoshi et.al. "Mining product reputations on the web", 2002 pp 341-349. Available at: https://dl.acm.org
[5] Pang Bo, Lee Lillian, Vaithyanathan Shivakumar, "Thumbs up? Sentiment classification using machine learning technique", 2002 pp 79-86 [Online]. Available at: https://dl.acm.org
[6] Wiebe M Janyce et.al. "Development and use of a gold-standard data set for subjectivity classification" 1999 pp 246-253 [Online]. Available at: https://dl.acm.org

[7] Liu Feifan, Wang Dong et.al. "Improving blog polarity classification via topic analysis and adaptive methods" 2010 pp 309-312 [Online]. Available at: https://dl.acm.org
[8] Medhat Walaa, Hassan Ahmed, Korashy Hoda, "Sentiment analysis algorithms and applications: A Survey" 2014, vol 5(4) pp. 1093-1113 [Online]. Available at: https://www.sciencedirect.com
[9] Miller A George et.al. "Introduction to Wordnet: An on-line lexical database" 1990 pp 235-244 [Online]. Available at: https://academic.oup.com
[10] Liu Jingjing, Cao Yunbo et.al. "Low quality product review detection in opinion summarization" 2007 [Online]. Available at: www.aclweb.org
[11] Miller Mahalia, Sathi Conal et.al. "Sentiment flow through hyperlink networks" 2011 [Online]. Available at: https://www.aaai.org
[12] Chamlertwat Wilas et.al. "Discovering Consumer Insight from Twitter via Sentiment analysis", 2012, Vol 18(8) pp 973-992 Available at: https://www.semanticscholar.org .
[13] O'Connor Brendan et.al. "From Tweets to Polls: Linking Text Sentiments to Public Opinion Time Series", 2012 pp 122-130 Available at: https://www.aaai.org .
[14] Ceron Andrea, Curini Luigi et.al. "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France", 2014 vol16(2) pp 340-358 Available at: http://nms.sagepub.com/content/16/2/340
[15] Yu Yang, Duan Wenjing et.al. "The impact of social and conventional media on firm equity value: A sentiment analysis approach", November 2013 vol 55(4) pp 919-926 Available at: https://www.sciencedirect.com
[16] Ortigosa Alvaro et.al. "Sentiment analysis in Facebook and its application to e-learning", February 2014, vol 31 pp 527-541 Available at: https://www.sciencedirect.com
[17] Greaves Felix et.al. "Use of Sentiment Analysis for Capturing Patient Experience from Free-Text Comments Posted Online", November 2013, vol 15(11) Available at: https://www.ncbi.nlm.nih.gov
[18] Poecze Flora, Ebster Claus et.al. "Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts", May 2018, vol 130 pp 660-666 Available at: https://www.sciencedirect.com
[19] Ray Paramita, Das Kumar Pranab et.al. "Demonetization and its aftermath: an analysis based on twitter sentiments", November 2017, vol 43 pp 186-196 Available at: https://www.sciencedirect.com