

## A Survey About Techniques On Opinion Mining And Sentiment Analysis Through Online Reviews

Dr.J.Chockalingam M.Sc., M.B.A., M.Phil., Ph.D.,  
M.Thirunavukkarasu M.Sc., M.Phil., (Ph.D.),

<sup>1</sup>Associate Professor/Research Supervisor PG & Research department of Computer Science Khadir Mohideen College, Adirampattinam-India

<sup>2</sup>Research Scholar PG & Research department of Computer Science Khadir Mohideen College, dirampattinam-India

Corresponding Author: Dr.J.Chockalingam

**Abstract:** Nowadays, the usage of internet world have been increasing as huge amount day by day. There are increasing in online way of analyzing the products and services through the feedback and the reviews collected from the users who used the products and availed the services. Analyzing the reviews is more important however obtaining the reviews is mandatory to take provisions. Here the most of the opinion analysis has been done over the reviews in the product or service platform itself. However the reviews collected and analyzed as much as well, the platform chosen makes diplomatic when it deals with valuable. The opinion analysis will be done in the social networks to be work with the real time reviews about the product and services. Thus the social networks reviews not only for analyzing the products and services. It can be used for the prediction of cyber-bullying messages and rumor detection in the social networks through the sentiment analysis performed by Natural Language Processing. This survey makes about the real time need for the sentiment analysis in the online social networks and its techniques used for the opinion mining.

**Key Terms:** Opinion Mining, Sentiment Analysis, Cyber-bullying Messages, Rumor Detection, Natural Language Processing

Date of Submission: 27-05-2018

Date of acceptance: 10-06-2018

### I. INTRODUCTION

Text Processing always plays major role in multimedia applications. The applications based on text processing such as rumor detection, review analysis, human stress detection and cyber-bullying message detection. In this scenario, the online social networks plays a major role in the analyzing of the products and reviews as well as some of the preventive measures for the social network can also be identified. The cyber-bullying messages in the social networks literally play normal or abnormal conditions in the people's environment. Thus the words needed to be avoided in that sentence that will make worse in the people's mind and nature. It will make some worse and initiate the unwanted things in the social human life.

Apart from the crucial things in social environment through the messages, the rumor makes worse in the environment. The rumors should be predicted as much as earlier through the online social networks text posts on their walls. The text processing plays major role in the rumor detection and avoidance in social networks.

Despite of all above, the most serious issue is to save the human value. The most of the people shared their entire feelings in the social networks. Now the tremendous work going to find the people who are in stress to avoid the suicide attempts. They are periodically counseled with the care when it should be found.

The Natural Language Processing techniques [2] are used to obtain the text processing. The study of linguistics and statistics can intersected in the field of NLP. The main goal of the NLP is automated finding of the semi-structured and unrealistic language that humans use. This NLP study stems application in assorted fields like summarization, semantic analysis, text processing and classification.

Online user-generated reviews are of immense realistic bring into play, because:

1. They have turn out to be an foreseeable component of decision making process of clients and users on product purchases, hotel and tour package bookings, etc.,
2. They as a group form a very low cost and competent feedback channel, which facilitates improving businesses to remain track of their reputations and to high progress the quality of their products and services.

In a work exposed in existing more directly related to our research direction, Mendoza et al [13], have concentrated at the difference and characteristics in propagation behavior of false rumors and true news on Facebook and Twitter [13]. Additionally, Friggeri et al. [14] and Jin et al. [15] have analyzed the cascade and propagation structures of rumors on social networks. Specifically, Jin et al. concentrated on the rumor propagation of Ebola pandemic virus and analyze the news spread and found that rumors can stretch just like true news. In all of these cases the properties of the actual entity that is being spread—be it a post or message, knowledge, or a virus— is never analyzed or taken into consideration in the models. In contrast, our exertion will be come across at the content of the messages or posts or tweets being spread in addition to the propagation behavior of these messages and any information that might be accessible about the persons involved in the rumor propagation.

In this paper, we discuss as follows: First the sentiment analysis and its needs and opinion mining overview are discussed in the upcoming chapter. Next to the major concept, the analysis techniques are discussed with the proven results obtained by the existing authors. The survey should be taken for the sentiment analysis in the online social networks through various existing works.

## **II. BACKGROUND KNOWLEDGE**

### *A. Cyberbullying Message Detection*

Cyberbullying [4] can be defined as destructive, most aggressive and intentional actions executed by an individual or a group of common people via online social networks such as posting messages or make tweets and posting comments against a victim. Different from conventional bullying that habitually occurs at social places during face-to-face communication, cyberbullying on online social media can come to pass anywhere at any time and serious and heavily spread over the people.

Rui Zhao and Kezhi Mao in [4] has defined the bullying message and its type of impacts. For serious bullies, they are easy to impair their end target feelings because they don't want to face someone and can hide behind the Internet. For impact victims, they are free to expose the harassment since all of people, especially youth, are persistently connected to Internet or online social media.

### *B. Rumor Detection*

Suchitha Jain et.al. a[17] discussed the details about the misinformation on twitter and how to detect the rumors spread in the social media. The propagation of rumors being distributed to large masses is becoming a massive problem nowadays. Since a huge number of online users are showing to the news instantly and they have a propensity to suppose whatever in sequence of data it has to be true, an environment of anxiety is developed.

For example, The Indian Government announced the demonetization of newly introduced 2000 rupees notes from the reserve bank of India. The government and appropriate authorities had to expend a lot of money and time to pull through from the wavering thus caused. Here the social networks are seriously concentrates for this type of misinformation spreading and the persons who spread the rumors.

### *C. Stress Detection*

Psychological stress should be made to the serious threat to the human's life. The people shared their mind feelings to the social media. The posts should be periodically identified the sentiments and seriousness that should identify the people who are in stressful condition. In [6], H Lin et.al. conveyed the importance the stress detection based on the assertion statements in the online social media. They proposed the framework based Convolutional Neural Networks. This is used for analysis because it having the capable of unified latent features learning from the multiple modalities and modeling the correlations based on factor graph model.

They analyzed the tweet streams about the assertion statements to obtain the psychological stress detection using the proposed framework.

## **III. TECHNIQUES USED**

To obtain the opinion mining for the above said purpose based applications are as following techniques.

### *A. Sentiment Analysis*

The sentiment analysis is used to define the characteristics of the sentence used in the social networks. Many sentiment analysis techniques have been developed for earlier period years to support the online the users in digesting the large amount of unstructured review data in social networks.

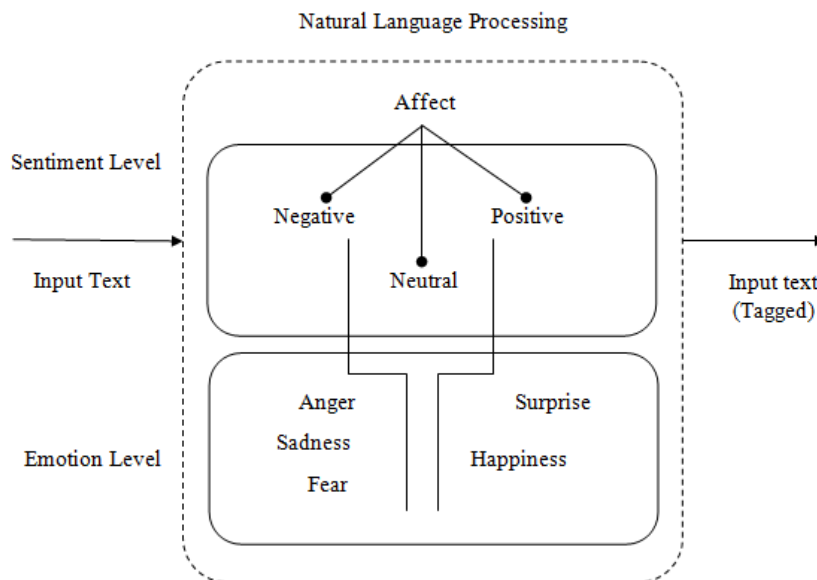


Figure 2.1. Framework for Sentiment Analysis

Sentiment analysis [16] is performed by using text classification approaches which are lexicon and machine learning based approaches. Lexicon based approach is obtained through the dictionary based approach and corpus based approach. Machine learning techniques in data mining for text processing are most widely used to classify and to predict the sentiment (text characteristics) as either positive or negative sentiment or may be in neutral sentence that haven't positive or negative. Machine learning algorithms are most important for classify as either supervised or unsupervised approach. Supervised approach can able to make process the labeled dataset where the training set has already assigned its nature on sentiment. Unsupervised approach can able to perform the process in unlabeled dataset where caption for the review is not defined with its label [2].

The sentiment analysis can also be obtained through the Latent Dirichlet Allocation was discussed by Zhen Hai et.al. [1] and the data mining classification techniques such as Support Vector Machine [9] [10], Naïve Bayes Classification and Decision Tree Induction [7]. The supervised machine learning models [7] [8] [11] such as Naïve Bayesian Classification and Support Vector Machine. In the above preceding said models, enhanced SVM performs better than the Naïve Bayesian Classification.

Support Vector Classification [12] is used to classify and obtain the prediction results through the classification and regression analysis in sentiment based analytic engines. In this paper, the discussions made for the overview of the sentiment analysis through the data mining techniques like SVM and also speak about the LDA Models [17]. There are lot of enhancements can be proposed to accomplish the new grievances in the data regression analysis.

### *B. Natural Language Processing*

The three class categorization problem (positive / neutral / negative) has most traditionally been used for sentiment analysis by the Natural Language Processing [2]. The sentences in the messages or posts should be analyzed after the text processing through this NLP. The NLP defines the right sentence content to the extent knowledge to classify the sentence behavior. Here it tries to address the sentiment classification problem with the reviewed and gathered set of text features with their content.

The prediction of emotional and sentence level affect in text obtained in the message or post or tweet is a topic that is mainly and closely related to NLP, but it has also fascinated the consideration of the TTS synthesis research community [18]. As far as we know, this work is one of the first attempts to adapt conventional Sentiment Analysis methods to the Text to Speech synthesis requirements.

Given that the sentence content information provided by a sentence obtained from the online social networks is rather condensed, some traditional approaches based on the Data Mining Machine Learning methods also developed using additional texts to infer further links with text emotional affect. Other innovative works, as an alternative, investigated into the relevant behavioral characteristics of the available input text of analysis without expanding the data with knowledge.

It allows the most expecting knowledge for the automatic language processing over the sentence knowledge learnt from the training data samples. The following process chain is used to obtain the natural language processing with sentiment knowledge.

#### C. Classification techniques

The data mining classification techniques are consistently used for the analysis of sentiment in the sentences posted in the social media as well as in the review platforms. The classification algorithms such as Decision Tree Induction, Naïve Bayes Classification and Support Vector Machine are used for the most cross domain and unstructured text data.

#### D. Latent Dirichlet Allocation Model

It is very complex to predict the accurate reasons at the back sentiment variations as number of posts or tweets are more than multiples for the target event. In the existing work proposed by Bholane Savita and Deipali Gore [11], they come across the sentiment polarity as well as causes following sentiment change for the target in contemplation using two LDA models. They think about two tweet sets i.e. foreground tweets and background tweets (noisy data). To take to mean the public sentiment variations, these background tweets should be removed. Tweets anthology in the variation period consists of the main reasons as well as the neutral tweets i.e. background tweets which had been discussed from a long time. The identified models are Foreground and Background LDA (FB-LDA) and Reason Candidate and Background LDA (RCB-LDA).

### IV. CONCLUSION

It is conclude that the social network post and messages are periodically need to classify based on the sentiments in the social networks. The different types of applications are used with the purpose of SA that is unpredicted with the accurate results. The most important problem is social network having the most number of domains in the text data based on the data mining classification techniques. The concentration should be needed on the data mining techniques so that the different domain based text data and huge in number of tweets are analyzed accurately.

### REFERENCES

- [1]. Zhen Hai, Gao Cong, Kuiyu Chang, Peng Cheng and Chunyan Miao, "Analyzing Sentiments in One Go: A Supervised Joint topic Modeling Approach", IEEE Trans. on Knowledge and Data Engineering, Vol.29, No.6, June 2017, pp.1172-1185.
- [2]. Nihar Ranjan, Kausal Mundada, Kunal Phaltane and Saim Ahmad, "A survey on Techniques in NLP", Intl. Journal of Computer Applications, Vol.134, No.8, January 2016.
- [3]. Hetl Bhavasar and Amit Ganatra, "Variations of Support Vector Machine Classification Technique:A Survey", Int. journal of Advanced Computer Research, Vol.2, No.4, December 2012.
- [4]. Rui Zhao and Kezhi Mao, "Cyberbullying Detection based on Semantic-Enhanced Marginalized Denoising Auto-Encoder", IEEE Transactions on Affective Computing, vol.8, no.3, September 2017, pp.328-339.
- [5]. Suchita Jain, Vanya Sharma and Rishabh Kaushal, "Towards Automated Real-Time Detection of Misinformation on Twitter", Intl Conf. on Advances in Computing, Communications and Informatics, Jaipur, India, November 2016.
- [6]. H Lin, J Jia, J Qiu, Y Zheng, L Xie, J Tang, L Feng and Tat-Seng Chua, "Detecting Stress Based on Social Interactions in Social Networks", Journal of LATEX class files, vol.13, no.9, September 2014.
- [7]. Han. J, Kamber.J, and Pei.M, "Data Mining: Concepts and Techniques", Elsevier 2<sup>nd</sup> Edition.
- [8]. Chih-Wei Hsu, Chih-Chug Chang and Chih-Jeh Lin, "A Practical Guide to Support Vector Machine Classification", Dept. of Computer Sci., National Taiwan Univ., Taipei, 106, Taiwan, <http://www.cse.ntu.edu.tw/~cjin> 2007.
- [9]. Fang Luo, Cheng Li and Zehui Cao, "Affective feature based sentiment analysis using SVM classifier", IEEE Int. Conf. on Computer Supported Cooperative Work in Design, May 2016.
- [10]. Bhumika.M and Vimalkumar.B, "Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis", Int. Journal of Computer Applications, Vol.146, No.13, July-2016.
- [11]. Bholane Savita and Deipali Gore, "Sentiment Analysis on Twitter Data using Support Vector Machine", Int. Journal of Computer Science Trends and Technology, Vol.4, No.3, Jun1 2016.
- [12]. Gaurangi P, Varsha G, Vedant K and Kalpana D, "Sentiment Analysis using Support Vector Machine", Int. Journal of Innovative Research in Computer and Communication Engineering, Vol.2, No.1. January 2014.
- [13]. M. Mendoza, B. Poblete, and C. Castillo. Twitter under crisis: Can we trust what we rt? In Proceedings of the first workshop on social media analytics, pages 71–79. ACM, 2010.

- [14]. A. Friggeri, L. A. Adamic, D. Eckles, and J. Cheng. Rumor cascades. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 2014.
- [15]. F. Jin, W. Wang, L. Zhao, E. Dougherty, Y. Cao, C.-T. Lu, and N. Ramakrishnan. Misinformation propagation in the age of twitter. *Computer*, 47(12):90–94, 2014. ACM, 2010.
- [16]. B.M.Jadav and V.B.Vaghela, “Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis”, *Intl. Journal of Computer Applications*, vol. 146, no.13, July 2016, pp.27-32.
- [17]. Shulong Tan, Yang Li, Huan sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen, Xiaofei He, “Interpreting the Public Sentiment Variations on twitter”, *IEEE Transactions on Knowledge and Data Engineering*, vol.6, no.1, September 2012.
- [18]. Alexandre Trilla and Francesc Alias, “Sentence-based Sentiment Analysis for Expressive Text-to-Speech”, *IEEE Transactions on Audio, Speech and Language Processing*, vol.21, no.2, February 2013, pp.223-233.

Dr.J.Chockalingam "A Survey About Techniques On Opinion Mining And Sentiment Analysis Through Online Reviews." *IOSR Journal of Engineering (IOSRJEN)*, vol. 08, no. 5, 2018, pp. 23-27.