A Survey on Sentiment Analysis Techniques on Social Media Data

Shah Shaheena Ashraf[1], Sahil Verma[2],Kavita[3]

[1]Student, M.Tech, EMGOI, Ambhala
[2]Assitant Professor, Dept. Of CSE, E-Max Group of Institutes, Ambhala
[3]Assitant Professor, Dept. Of CSE, E-Max Group of Institutes, Ambhala

***Abstract*-Trend of social networking sites is at peak, so ocean of data is being created and generated. Micro blogging sites are used by billions of individuals to express their opinions on different topics related to different fields .Micro blogging has become a dynamic platform of communication between Internet users. Twitter a micro blogging service, has become the main portal where users prompt their views, emotions and opinions. Twitter is the fundamental social networking website. Opinion mining and sentiment analysis is most trending hot topic with numerous multi applications ranging from simple to complex. With the speedy improvement and progress of social network websites, sentiment analyses have become the main central talk of every conversation. This paper presents a review of various research works done in this field.**

Index term: **Social media, Twitter, opinion mining, sentiment analysis**

I. INTRODUCTION

Emotions and opinions are essentially present in all the individuals. World cannot be visualized without them. All humans' engagements go round the emotions, feelings and opinions. They influence the individual's life by the way we think, what we do and how we act. Access to internet or the information in the social media is currently no longer a subject of concern, as there are gigabytes of new information being generated every day that is accessible to one and all individual. Frankly speaking it has changed the approaches and techniques of information sharing. The users of internet not only consume the existing content on web, but instead interpret this available data and produce new fragments of information. Nowadays people also share their ideas, feelings, opinions and knowledge with the whole world rather than only commenting on the existing information, events, bookmark pages and providing feedbacks.

In this fashion, besides being a reader, whole community is transformed into writer. The current prevailing platforms are Blogs, Forums, Wikis and social networks where the customers and users access the information present in the web,provide opinions and receive feedback from other users on various topics ranging from agriculture to industries, education to health,politics to travelling ,customer reviews to product ratings etc. Presently many scholars and researchers are excelling in this field. They are trying to extract opinion information to explore and summarize the opinions communicated automatically and spontaneously with computers.This new field of research is nowadays termed as Opinion Mining and sentiment analysis. Researchers have developed numerous techniques until now which are solution to the many problems. Nowadays Information Retrieval (IR) and Natural Language Processing (NPL) are the fields where Opinion Mining and Sentiment Analysis are studied**.** The promotion and acceleration of web technology has lead to huge volume of data for users in the web and creation of new data to**o.**

SOCIAL MEDIA:

Social media is used by every fifth individual for one purpose or the other. Among them are content video sharing sites, blogging and micro blogging sites, friendship networks, e-commerce, e-banking etc. Progressively participation and engrossment of users on the internet is growing tremendously. Among such contributions is feedback of users in social networking sites. The recent inclination allows the users to take healthier decisions regarding a specific service or purchase about a particular product. Reputation and desirability of the product is also verified and checked by this. It also helps in extracting the positive and negative characteristics of the product by reading the reviews and feedbacks of different users. Currently many social media platforms has become the dynamic part of human's life. Twitter, Facebook, YouTube, Google Plus, LinkedIn, WhatsApp are among the most prominent and widespread platforms.

TWITTER

The most renowed and known micro blogging service is twitter, where the messages are read and posted by the user which are 148 characters in length. Twitter messages are also known as Tweets.Twitter is a important micro blogging platform which has discovered what is occurring across the globe at any instant of time.

TWITTER ANALYSIS

In twitter analysis tweets are treated as raw data. Different methods and algorithms are used that spontaneously refine tweets into positive, negative and neutral sentiments. Sentiment analysis is used in prediction democratic electoral events, stock market, consumer brands, sports movie collection at box office, celebrities etc.

II. SENTIMENT ANALYSIS

A process that systematizes withdrawal of approaches , opinions, assessments, visions, views, feelings ,emotions, sensations, excitements, attitudes from verbal(text), nonverbal(speech),tweets and database sources through Natural Language Processing (NLP). Sentiment analysis comprises of organizing opinions in text into classes like "positive" or "negative" or "neutral".It's also termed as subjectivity analysis, opinion mining, and appraisal extraction.

Natural language processing (NPL) and information extraction task leads to sentiment analysis. Writer's opinionsstated in comments or reviews are extracted by this method. Besides extracting the polarity from the text sentiment analysis also extracts features from the text.

Sentiment analysis applications can be felt worldwide. Applications opinion mining can be seen in business and government intelligence. One of the central factors behind schedule corporate interest is the business intelligence in the area. With the unexpected low sales computer manufacturers feels great disappointment, confronting themselves with several questions such as "why are the sales low?" while the concreate data was quite relevant, it require users reviews and feedback on a particular product characteristics.

III. RELATED WORK

Ortigosa and Alvaro [1] proposed a new procedure for sentiment analysis in social sites like gaintFacebook, twitter etc., it starts when users write messenges, it helps in: (i) Determining the sentiment polarity which may be positive, negative or neutral of social network users, which is revealed from the posts of users; and (ii) normal sentiment polarity of users is modeled and majoremotionalfluctuations of users are explored. Author has implemented this method in Sent Buk, a Facebook application also included in the paper. It is obligatory for a system to fetch and store this data information so as to take conclusions based on information, opinions and emotions of the users. Details and statistics about the users' opinion can also be fetched by the most significant method by directly asking them to fill questionnaires through which their opinion can be considered. Though, for a some user this job can be too inefficient, slow and tiresome.

Pak and Alexander proposed [2], Micro blogging is becoming a very popular communication and knowledge sharing tool among Internet users globally. Billions of users exchange their knowledge on different aspects of life each day. Therefore these micro blogging sites are full resources of information and data for sentiment analysis and opinion mining. Because micro blogging is recent trend, there are some research works that were devoted. The main focus is on using Twitter which is the most widespread micro blogging website, for the sentiment analysis, in this paper. This paper indicates how to spontaneously obtain a corpus for opinion mining and sentiment analysis purposes. Then it performs standard linguistic analysis technique of the collected corpus and explains acquired phenomena. By using the corpus, Author builds a sentiment classifier, which is capable of determining positive, neutral and negative sentiments for the whole document. Experimental conclusionsindicate that the proposed techniques are more efficient and perform better as compared to previously proposed techniques. Author worked with English; however, this proposed technique can also be used with other language.

Agarwal and Apoorv [3] explained one such prevailing micro blog named as Twitter and construct models to categorize the "tweets" into positive and negative sentiment or they can be neutral. Author build novel models for two classification: first one is a binary task of classifying sentiment of users into positive and negative classes and second is a 3-way task of classifying sentiment of users into positive, negative and neutral. Author experiment with two kinds of models: (1) unigram model which is based on feature modeling (2) a tree 30 kernel based model. The paper focuses on features of existing proposed technique in past literature and proposes novel technique for the feature based model. Author design a new representation for tweets, for the tree kernel based model. Author use a unigram model, which work better for sentiment analysis for Twitter data in the past. Result from paper indicates that a unigram model is really a hard baseline [4]. Feature based model that used 100 features gives similar accuracy as compared to the unigram model that used about 10,000 features. Tree kernel based model gives improvement outperforming both these models by a significant margin.

Aisopos and Fotis [4] presented with Microblog content, some serious challenges are associated. Few of them are the applicability of sentiment analysis used in past and different classification methods caused by their inherent characteristics of content. To resolve them, author introduces a technique that depend on on two orthogonal and matching sources of evidence: context-based method taken by polarity ratio and content-based features acquired by n-gram graphs. Both the methods are language-neutral and tolerant to noise; guarantee

high robustness and effectiveness in the manner author are considering. To ensure this methodology can be applied into concrete applications with huge amount of data, aim should be improving its time efficiency. Thus author propose substitute sets of features having minimum extraction cost, explore dimensionality reduction techniques and discretization techniques and also research with numerous different classification algorithms.

Balahur and Alexandr[5],Current time belongs to sentiment analysis as there is remarkable progress in the field of sentiment analysis, especially in subjective text types (like movie or product reviews). During publishing the target is flawless and uniquely stated through the text distinguishes these subjective texts. We comprehended that opinion mining is different from that of other text types, by following diverse annotation efforts and the examination of the issuesmet. We recognized three subtasks that must to be addressed: defining the target; separating the bad and good news content from the bad and good sentiment expressed; and finally analysis of clearly mentioned opinion that is expressed unambiguously, not needing understanding or the utilization of world knowledge. Additionally, we distinguish three dissimilar views on newspaper articles (text, author and reader), which have to be handle differently while analyzing sentiment.

Horakova and Marketa[6] present a model whish provide a view of business intelligence by gathering the tweets or messages from social networking sites. In this framework two layers of sentiment analysis were found, the data processing layer and sentiment analysis layer. Data processing layer is conserned with data collection and data mining, while sentiment analysis layer use an application to present the result of data mining.

Adarash MJ and Pushpa Ravi Kumar [7] have seen the impression of Micro Blogging site Twitter on recent trends and problems. The sentiment analysis on Twitter helps to determine the behavior of usersand to study the commercial. Twitter data is beinganalyzed in several ways, the occurrence of words like good, bad and also presence of emoticons in the tweets can be used to conclude the sentiment. Based on the followers and the followees the Twitter users can be categorized into positive, negative and neutral users and their behaviors can be calculated from their tweeting and retweeting activity.Whether a tweet is sent by a human or a bot can be checked on the basis of sentiments. Inadequate datasets and other rational issues are taken into thought for obtaining the results. If adequate and larger datasets containing all aspects into consideration for opinion mining are used, this will certainly lead to some considerable results which can be used for forthcoming research in social networking.

Varsha Sahayak Vijaya Shete Apashabi Pathan [8] explained that the twitter is a ultimate micro blogging service which is used to identify what is trending across the globe at any instant of time. In the paper, they have concluded that sentiment prediction in twiter can be fetched from social media features. They have used three machine learning algorithms which will contribute to beat three models namely unigram, feature based model and tree kernel model by using Weka. So, their proposed system determines the sentiments of tweets which are take out from twitter. With the nuance and complication of opinions conveyed the difficulty amplified. Product reviews, feedbacks,etc. are relatively easy. Books, cinemas, graphic arts, music, composition are more difficult. They also implement features like capitalization/internationalization, emoticons, negation handling neutralization, and as they have recently become a integral part of the internet.

IV. CONCLUSION

A systematic review of all the work done in this field is presented and it was found that mostly methodologies for the classification of sentiments was developed in this thesis for educational data mining crisis in Indian market. Twitter API was used for streaming of tweets. Stemming was done to all words in order to extract the root words. TF-IDF score based approach was utilized and the score was calculated for each tweets. Feature Selection was applied on it using Chi Square method and information gain. The extracted features form a term document matrix which is utilized in the classification algorithm. Classification algorithms can be used for this purpose in future.

REFERENCES

1. Ortigosa, Alvaro, José M. Martín, and Rosa M. Carro."Sentiment analysis in Facebook and its application to e-learning." *Computers in Human Behavior* 31 (2014): 527-541.
2. Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." *LREC*. Vol. 10. 2010.
3. Agarwal, Apoorv, et al. "Sentiment analysis of twitter data." *Proceedings of the Workshop on Languages in Social Media*. Association for Computational Linguistics, 2011.
4. Aisopos, Fotis, et al. "Content vs. context for sentiment analysis: a comparative analysis over microblogs." *Proceedings of the 23rd ACM conference on Hypertext and social media*. ACM, 2012.
5. Balahur, Alexandra, et al. "Sentiment analysis in the news." *arXiv preprint arXiv:1309.6202* (2013).
6. Horakova, Marketa. "Sentiment Analysis Tool using Machine Learning." *Global Journal on Technology* (2015).

7. Adarash MJ and Pushpa Ravi Kumar" Survey: Twitter data analysis using opinion mining" *International Journal of Computer Applications* Vol. 128- No. 5, October 2015.
8. Varsha Sahayak, Vijaya Shete and Apashabi Pathan. "Sentiment Analysis on Twitter Data" *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*Issue 1, Volume 2 (January 2015).

