Opinion and Sentiment Variations Analyzed on Twitter and Facebook

Shiramshetty Gouthami
Research Scholar,
Osmania University,
Hyderabad,India.

Dr. Nagaratna P Hegde
Professor,
Dept. of CSE, VCE, Ibrahimbagh,
Hyderabad, India.

Abstract: Twitter platform is priceless to comply with the public sentiments. Realizing customers factor of views and reasons in the back of them at more than a few point is an principal learn to take detailed decisions. Categorization of positive and poor opinions is a procedure of sentiment analysis. In this paper, we have carried out further analysis to hold the useful insights for PSV utilizing tweets and facebook feedback with a exact target. We propose a Novel reason Mining Algorithm to seek out the feasible explanations affecting PSV in huge sentiment variation interval. It uses incremental iterative technique to refine the final record of most influential factors. To provide more intuitive representation, the algorithm ranks a collection of mined motive candidates. The purpose candidate with probably the most quantity of tweets/feedback is the main intent for sentiment variation in that point period. The quality of mined causes most commonly is dependent upon the high-quality of clusters formed previous to reason mining.

Key Words: Public Sentiment Variation, Novel Reason Mining Algorithm, Twitter, Facebook.

I. INTRODUCTION

Introduction Sentiment evaluation is often referred to as opinion mining refers to the use of regular language processing goals to check the perspective of a speaker or a writer with appreciate to a few topic. The angle is also his or her judgment or analysis [1]. The upward thrust of social media indicative of blogs and social networks has pushed interest in sentiment analysis. Because of the proliferation of stories, ratings and other forms of online opinion ,on-line expressions has become a form of platform for companies viewing to market their commodities, establish new possibilities and manipulate their reputations[2]. Sentiment analysis is an unique new research discipline with the skills for a number of actual world purposes the place found out opinion information can be utilized to

help individuals or organizations or organizations to make better decisions. For illustration, if public sentiment changes broadly on some products, the related corporations may just need to comprehend why their merchandise receive such feedback. If terrible sentiment toward Barack Obama raises vastly, the White condo Administration place of business is also eager to know why men and women have converted their opinion after which react thus to reverse this development [2]. As primary application of sentiment analysis is to classify a given text to one or more pre-outlined sentiment categories and can be used for determination making in more than a few domains. It is normally elaborate to find the distinct factors of sentiment versions due to the fact they may involve elaborate inside and external reasons. It is determined that the emerging issues discussed within the version period might be highly regarding the actual explanations behind the variations. When persons state their opinions, they more commonly state factors (e.g. Some targeted routine or subject matters) that support their current view [2].The Proposed process can analyze public sentiment variations on Twitter and mine possible causes at the back of such variations. To track public sentiment, The naïve bayes algorithm is used to receive sentiment data towards interested targets (e.g. Obama, Apple) in each tweet. After obtaining the sentiment label for each tweet, For monitoring the public sentiment concerning the corresponding goal some descriptive data (e.g., Sentiment percentage) is used. In the monitoring curves huge sentiment variants may also be detected with a pre-outlined threshold (e.g., the percentage of bad tweets increases for more than 50%).

II. RELATED WORKS

Sentiment analysis is the process of analyzing the opinions which are extracted from different sources like the comments given on forums, reviews about products, various policies and the topics mostly associated with social networking sites and tweets. Pang et al. [2] work on the supervised machine learning methods existing for analyzing sentiments.

Advantages: Different machine learning methods reduce the structural risks.

Disadvantages: This method is not able to analyze possible reasons occurred behind the public sentiments. The Supervised machine learning methods demands for large amount of labeled training data which is expensive. It may not work when training data are insufficient.

M.Hu and B. Liu [3] works on novel techniques which are useful to collecting and summarizing customer reviews. It identifies customer opinions and after that decides whether it is categorize into positive or negative.

Advantages: It predicts movie sales as well as elections which help for making decisions. It also provides a featurebased summary for large amount of customer opinions.

Disadvantages: It will not calculate the opinions strength and also not represent opinions with its verbs, adverbs and nouns.

W.zhang et al. [4] studied of opinions retrieval from blogs. In this paper, they have presented a three-component opinion retrieval algorithm.

1. Selects information retrieval module.

2. Classifies document into optionative document.

3. Rank document in certain order.

Advantages: It gives higher performance than state-of-art opinion retrieval methods.

Disadvantages: It is not able to handle more general writings and crossing domains. It also not selects detail features.

III. PROPOSED METHOD

The steps performed for inspecting public sentiment versions and mining the valuable insights behind these editions are described in this part. First, the tweets and facebook comments concerning precise target (e.g. "Narendra Modi")
are extracted from the collected tweets and comments. The extracted tweets and feedback are pre-processed to do away with the noise in the data. The mission of preprocessing is important to make the sentiment analysis extra effective. Then these tweets/comments are categorised into three courses like optimistic, bad and impartial via making use of Stanford NLP Classifier. On the groundwork of classification of each tweet and

remark, the sentiment variations regarding the distinctive objectives have been tracked with the support of some descriptive records. Subsequently, the reasons at the back of these public sentiment versions have been analyzed through making use of the novel motive Mining Algorithm.

All the tweets and comments regarding the special goal are extracted by means of filtering the tweets/comments which incorporate the key words of the goal. Frequently the noise within the tweets and comments influences the results of sentiment analysis. The preprocessing approaches get rid of the noise and make them efficient for sentiment classification. The preprocessing strategies used in this work are described right here within the following subsections.

A. Stop Word Removal
Extremely common words are not considered in order to speed up results. These filtered words are known as 'Stop Words'. A list of commonly used stop words is retained and it is used to remove all the stop words from tweets/comments.

B. Stemming
Stemming is done for reducing inflected (or sometimes derived) words to their stem, base or root form. For example the word "Happiness", indicate the positive sentiment. To extract the correct sentiment from these type of words. They
should be stemmed to its root i.e. the word "Happy". A popular Porter stemmer algorithm is used for stemming process.

C. Conversion of Slang Words
User Tweets and comments often contain slang words. The slang words like omg, lol play important role in sentiment analysis. A list of commonly used slang words is maintained along with their translation in standard form. The list is
created by using Internet Slang Word Dictionary. First we converted the slang words into their standard forms and then used them in tweets/comments to analyze the sentiment.

D. Removal of URL
Most of the tweets and comments contain URLs. These URLs should be removed to make the task of sentiment analysis more effective. To remove the URLs a regular expression containing a pattern for URL is used. The string matching with the pattern is removed from the tweets/comment.
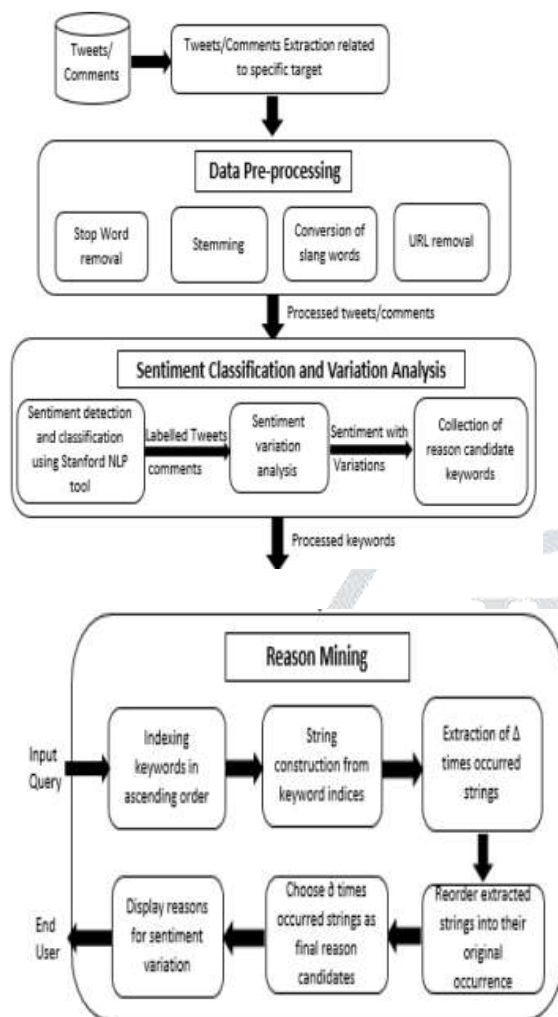
Fig 1: Methodology Flowchart

To categorise the tweets and feedback as confident, bad and neutral, a sentiment analysis device of Stanford NLP framework is used. This tool is established on precept of maximum Entropy Classifier, which classifies the tweets/comments into positive, poor and neutral lessons. This Stanford NLP software has verified its exceptional and accuracy, as various sentiment processing applications has adopted it as a widespread sentiment analysis instrument in computer studying. Accordingly we can estimate the chance that an opinion could include positive, poor or impartial sentiments.

After classifying all of the extracted tweets and feedback about a target, we analyzed the sentiment variants using the number of optimistic and poor tweets/feedback. But the number of constructive and negative tweets will not be so useful, as this quantity may alternate consistently. On this work, we regarded percent of constructive, negative and neutral tweets/comments to analyze the sentiment variations, e.g. The percent of

terrible/optimistic tweets/feedback is growing greater than 50%. The facts bought after classification is the priceless indicator for sentiment variant over the period of time. The next figures show the general public sentiment variants closer to "Narendra Modi" on Twitter and fb from January, 2015 to April, 2015.
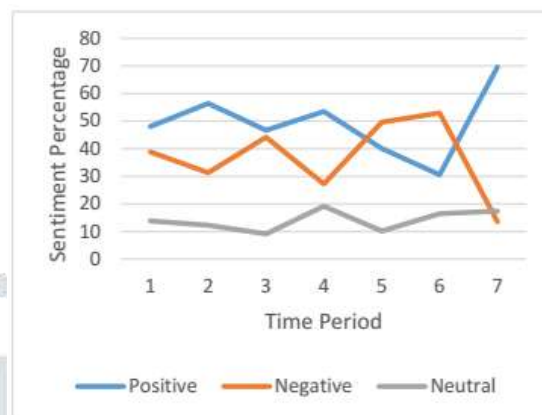

Fig 2: Public Sentiment Variation towards "Narendra Modi" on Twitter
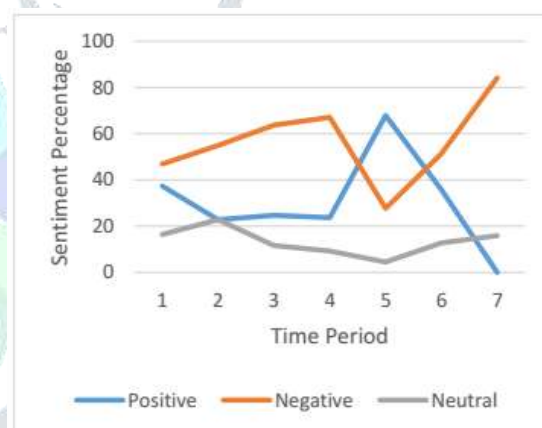

Fig 3: Public Sentiment Variation towards "Narendra Modi" on Facebook

Motives are the core part of tweet/remark, which is simply dominating factor of our present context. In reality finding reasons is not an easy challenge, however we will find clues by using inspecting the tweets/comments in variant period. Folks give their opinion founded on current cause for special entity e.g. Character, institution, event, and so on. A Novel reason Mining Algorithm is carried out to specify record of factors within particular time interval for special entity. Earlier than applying this algorithm, the entire keyword phrases (distinguished words) of a tweet/remark were individually saved in a desk to make processing easier. A separate list of the most usual came about phrases for the period of specific time interval is created. These most almost always passed off words are sub-ingredients of our primary motives to be mined. Then the association

relationship between these most most of the time happened phrases and their combined incidence rely is decided. This mixed prevalence count will gave us sorted record of predominant motives.

We used following notations in algorithm to represent main entities to be considered in calculations.

W: time window of specific days
T: a set of user tweets within time window W
Δ: threshold value to be used to consider association occurrence count between keywords
∂: min value to be used to consider sequence occurrence count between keywords

Algorithm:

1. A set of user Tweets within the time window 'W' are -extracted from database.
2. In processing stage we already stored all the keywords of tweets in separate table. We take most occurred keywords within current time window 'W' and store in separate array in ascending order.
3. A string is constructed from array index of keywords.

For example, consider the following array of most commonly occurred keywords within specific window [government, Dubai, India, minister, prime, Indian, world, people, vote]

A sample tweet is shown below:
Modi IN Dubai. It's Respectable Modi that can changed the Indian scenario toward world power.
This tweet contains three words from above array of keywords (Dubai, Indian, people). Then based on their indices "1-5-7" string is constructed.

4. A string of array indx is finally considered if more than Δ (threshold value) number of tweets produces such string.
5. A set of string of array index is then sorted in descending order based on their occurrence counts.
6. This set of string of array index is again processed to construct string of keywords which is final reason.
7. Actually this reason is a sorted string. So to achieve exact ordered reason based on its exact sequence again it is compared with tweets.
8. ∂ (min value) times occurred word sequence is finally shown as mined reason

IV. CONCLUSION

This paper focuses on inspecting the general public sentiment variants and mining possible causes behind these variations on Twitter and facebook. To track the public sentiment, the Stanford NLP's sentiment evaluation device is used, which classifies the tweets/posts into confident, poor and impartial classes. On the foundation of classification, the general public sentiment variants had been tracked with the assistance of some descriptive statistics. The radical Reason Mining Algorithm can to find the feasible explanations at the back of public sentiment variants. It makes use of incremental iterative strategy to refine the final list of most influential causes. To provide extra intuitive illustration, the set of mined reasons are ranked in keeping with the number of tweets/comments associated with the unique intent. The reason candidate with essentially the most number of tweets/feedback is the important motive for the sentiment variation in that point period.

REFERENCES

[1] B. Pang and L. Lee, Opinion mining and sentiment analysis, Found. Trends Inform. Retrieval,vol. 2, no. (12), pp 1135, 2008.

[2] B. O Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, From tweets to polls: Linking text sentiment to public opinion time series, in Proc. 4th Int. AAAI Conf. Weblogs Social Media, Washington, DC, USA, 2010.

[3] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, XifengYan, Jiajun Bu, , Chun Chen, , and Xiaofei He,IEEE Transactions on "Interpreting the Public Sentiment Variations on Twitter," VOL. 26, NO. 5, MAY 2014.

[4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp. 79-86.

[5] Pang B., Lee, L., "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts",Proc. Of 42nd annual meeting of the association for Computational linguistics(ACL), 2004, 271-278

[6] Boiy, E., Hens, P., Deschacht, K., and Moens, M.F. "Automatic sentiment analysis in on-line text"

Proceedings Of 11th International Conference on Electronic Publishing(Vienna, Ausrtria) 2007

[7] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the Association for Computational Linguistics (ACL), 2002, pp. 417–424.

[8] M. Hu and B. Liu, "Mining and summarizing customer reviews," Proceedings of the tenth ACM international conference on Knowledge discovery and data mining, Seattle, 2004, pp. 168-177.

[9] Godbole,N., Srinivasaiah, M.Skiena S. "Large scale sentiment analysis for news and blogs", Proceedings of the International Conference on Weblogs and Social media (ICWSm'07), 2007

[10] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval 2(1-2), 2008, pp. 1–115.

[11] Anuj Sharma, Shubhamoy Dey "Artificial Neural Network Based Approach for Sentiment Analysis of Opinionated Text", ACM transaction, 2012

[12] A. Mudinas, D. Zhang, M. Levene, "Combining lexicon and learning based approaches for concept level sentiment analysis", Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York,NY, USA, Article 5, pp. 1-8, 2012

First Author: Shiramshetty Gouthami is presently working as Assoc.Prof in Department of Computer Science Engineering at Jayamukhi Institute of Technological Sciences,Warangal for the past 11 years. Presently, Ph.D scholar at Osmania University(T.S), India.

Second Author: Nagaratna P.Hegde, is currently working as Professor at Vasavi College of Engineering, Hyderabad.She has chaired various conferences and is actively trying to propagate Computer Science through her works. She has obtained her Doctorate from JNTU Hyderabad.Having 22 years of experience in teaching Computer Science both at undergraduate and post graduate levels, she is presently guiding eight members for their Ph.D. Her research interests include image processing; natural language processing, speech processing, and knowledge based systems, data mining and Big Data analytics and has to her credit around 45 papers published in journals and conference proceedings.