Survey on Location based sentiment analysis of Twitter data

[1] I.Karthika, [2] S.Priyadharshini
[1] Assistant Professor, [2] PG Scholar
[1] Department of Computer science and Engineering,
[1] M.Kumarasamy college of Engineering, Karur, India

Abstract --- **Big data is a concept used for collecting, storing, and analyzing large volume of data and provides decision making and also support optimization processes. Social media plays an important role in taking decisions about any products based on the reviews provided by the user. It accurately tells about the exact opinion of the user regarding the product. Twitter is one of the famous social network which is used by most the online user. The account holders are allowed to post their opinions in the form of tweets. In the previous approaches, the product reviews alone considered. In the proposed system, tweets of the products are analyzed based on the location and comparisons on the feature specific and gender specific categories. Also, the text tweets along with emojis and images could be analyzed. Tamil tweets can be considered for the analysis purpose.**

Index terms --- **Social media, Twitter, Opinion mining, emojis**

I. INTRODUCTION

Sentiment analysis is a technique or methodology which categorizes the products based on the reviews made on that product in any social network. A number of organizations in the olden days, used the paper based approach in order to collect feedback from the user. But this is not very efficient and hence they concentrate on the online feedback provided on the twitter. This helps the companies in analyzing the product and can extend their marketing expenditures where the sentiment is very low.

Based on the public comments provided by the customers, we can predict the nature of the product features, where the product lags and either it is successful or failure one. There are four steps involved in the analysis.

Data extraction phase involves the collection of tweets according to the keyword and the location given. The next phase is the data preprocessing which involves the filtering of the tweets with proper grammatical relation. Sentiment score phase involves the scoring process. Once analysis process is completed, the comparisons are made based on location, feature and gender.

II. LITERATURE REVIEW

Syed akib anwar [1] proposed that Public sentiments are the main things to be noticed for collecting the feedback of the product. It can be done by using sentiment analysis. The twitter is the social media used in this paper for collecting the reviews about any product. The reviews collected are analyzed based on the locations, features and gender. There are four steps involved in the paper: Data extraction which involves collecting the twitter data, data processing involves filtering out the redundant tweets and non grammatical relations, implementation involving the product analysis using sentiment score and result involves comparison between gender, feature and locations.

Xing Fang [2] discusses that Sentiment analysis is a technique used for categorization of the product based on the reviews of the user. The categories of the product are good, bad or neutral. In this paper, the general problem of the sentiment polarity categorization has been resolved. The sentiment polarity categorization consists of two phases: sentence level categorization and review level categorization. The sentence level categorization reveals the positive or the negative sentiments whereas the review level categorization reveals the star scaled ratings.

Carolina Bigonha [3] in this paper the social media or the social network site is a powerful tool in collecting the user data (ie.,) their opinion about any product or service can be found. One of the major site used by the people is the twitter. But analyzing the tweets produced by the users of the product is expensive. Also, the irrelevant tweets are eliminated for the analysis. Hence, a new methodology is proposed in order to extract the most influential users. To identify the most influential users two phases are involved: pre-processing and metrics analysis. In the pre-processing phase, the keyword of the topic based on which analysis needs to be done can be selected. Then, the related tweets are filtered out and the sentiment analysis is performed. Tweets collected are used for obtaining the user. In the metrics analysis phase, the network, polarity and quality values are calculated and they are combined together to form a single factor.

Petra Kralji Novak [4] This paper proposes that Emoticons shortly known as the emojis reveals the users emotions or their opinion about anything that they need to be expressed in the social networking sites. Emojis are used as short hand expressions that carry the emotional contents. To know about the emotional contents, the emoji sentiment lexicon named emoji sentiment ranking is used. It contains 751 most frequently used emojis for sentiment mapping process. Sentiment labels can be either,

positive, negative and neutral. The overall sentiment score of the tweets are calculated based on their position. The position can be determined by using the sentiment score s and its neutrality p_0 of the emojis.

Hailin Jin [5] proposes that Sentiment analysis is technique used for analyzing the social media analytics. It is mainly used for text information analysis. The opinions on any products are done based on the reviews provided by the users. Recently social media users are very much involved in usage of the images and videos rather than text messages. Hence sentiment analysis on such visual content helps in analyzing the exact user sentiments. Since the sentiment analysis on such visual contents is very tedious, a new methodology is employed in order to do image analysis. In this paper, Convolutional Neutral Networks(CNN) is used for this purpose. There are two approaches involved in this process: Visual sentiment analysis with regular CNN and Visual sentiment analysis with progressive CNN. The result is that Sentiment analysis is quite useful for data analytics tasks which can be used for prediction and forecasting.

Apoorv Agarwal and Owen Rambow [6] This paper proposes a pattern to filter out the opinions from the real time microblogging service, Twitter. A novel approach is used which combines the concepts of corpus based methods in order to identify the semantic of the tweets in the Twitter. It also uses the concepts of Natural Language Processing(NLP) and Machine Learning(ML) algorithms along with above two methods. The corpus based method is used to identify the semantic orientation of adjectives whereas the dictionary based one is used to identify the semantic of the verbs and adverbs. The overall sentiment can be calculated by using some linear equation that consolidates emotion identifiers. To calculate the sentiment score, the average of the strength of all opinion indicators are taken into account.

Therasa Wilson [7] utilizes the linguistic features for detecting the sentiment of twitter messages. There are three types of datasets collected in order to identify the opinion about the products: Hash tags dataset, Emoticon datasets and iSieve datasets. Hash tagged dataset consists of the collection of tweets with hash tags. It filters out the duplicate tweets, non-English tweets and also the tweets without hash tags. The tweets with the hash tags are used to select those tweets which can be used for development and training processes. Emoticons dataset contains the tweet set with both positive and negative emoticons. The data with both positive and negative emoticons are eliminated from the set. iSieve datasets denote the set of tweets that were collected on a selected specific topic which reflects the sentiments towards the topic. The next step is the preprocessing of the collected tweets which involves the three processes: tokenization, normalization and parts-of-speech tagging. The results show that sentiments are exactly identified but the benefits of emoticon training set has been minimized.

James Spencer and Gulden Uchyigit [8] This paper proposes a tool, Sentimentor which is used for the analysis of opinions on the data from twitter. Sentimentor tool uses the naive Bayes classifier in order to classify the tweets into positive, negative and objective sets. A sentiment classifier is built in order to determine the positive, negative and objective sentiments by using the above corpus. Sentimentor contains an interface which helps the users for analyzing the word distributions. It presents the analysis result in a pictorial format. This format helps the users to know the results of the analysis in a better way.

Aarathi Patil [9] This paper proposes that the sentiment analysis can be done on any product or event by using the social media based on the location. This helps in improving the product marketing expenditure based on the score produced for that product. There are four major steps involved in this paper. First step is to create Twitter application which is used to mine the twitter4j and analyze the data. Next, the tweets are collected by using the secret tokens from the twitter. These collected tweets are saved in an excel file. Preprocessing of data which are collected has been carried out. Once the preprocessing step is completed, the filtered out tweets are classified by using the Naive Bayes classifier. The sentiment scores are provided as 1 for negative sentiment, 2 for neutral sentiment and 3 for positive sentiments. The result is that the location based product opinions are collected by using the social media posts.

Vasavi Gajarla and Aditi Gupta [10] proposed that the emotions said by the images can be learned by prediction. The automatic tag prediction method is used for prediction of emotions from the images that are uploaded or posted at the social network sites. The emotions are categorized as 6 types: Happiness, Sadness, Fear, Disgust, Anger and Surprise. The twitter data is collected by using two steps: Querying of the image metadata and downloading the images from the server. Once the data collection process is completed, the images are analyzed for emotions. The pertained neural network is used to get the feature representation of the dataset and by using this; the data can be classified through one vs. all SVM classifiers. Here, the pre-trained model is the VGG-ImageNet. Next, Sentiment analysis is performed in order to classify the images according to their corresponding categories.

Kaushika and Uma [11] In this paper, a methodology named Word Sense Disambiguation is proposed for sentiment analysis of bilingual tweets. Word Sense Disambiguation is the ability to circumstantiate a word that has different senses which are used based on their usage pattern. In order to identify the sentiments in English and Tamil tweets, two phases are involved: In the first phase, path length similarity between the two synsets is computed with the help of WordNet. The translation of Tamil tweets involves a separate step. Once the tweets are collected by using Twitter API, the Tamil and English tweets are separated by using Tamil Corpus. The Tamil tweets are converted into English tweets by using Google translator. After the conversion of tweets, these are preprocessed for filtering out the efficient tweets. The Path Length Similarity is done for Word Sense Disambiguation. The Second phase consists of Support Vector Machine which can be used for classification of the tweets.

Gargi Mishra and Shivani Varshney [12] proposes an efficient methodology to determine the people opinions from real time twitter data. Two phases are mainly involved in this paper: In the first phase, the web crawler is used for extracting the real time data by ASP.net. The creped knowledge from the twitter is stored in the database. The procedure of the web crawler is as follows. The user enters the keyword and clicks for submit button. The crawler collects the tweets related to the keyword and stores them at the database. The second phase involves the process of analysis of collected tweets that are stored at the database. After preprocessing process, the classifier is used in order to classify positive, negative and neutral sentiments for which Support Vector Machine is the classifier. Thus, we can select a particular location for gathering the data related to that area in the form of pie chart.

Jianbo Yuan and Jiebo Luo [13] In this paper, Sentiment analysis is done on the images. But the analysis of images is much more difficult task to be processed. Hence, a methodology is used to ease of the image analysis (ie.,) Sentribute framework. The concept behind this framework is that the scene descriptor low-level features are extracted and these features are used to train the classifiers for generating 102 predefined mid-level attributes. These attributes are used for predicting the sentiments. In the case of the facial sentiments, the eigenfaces are used. This methodology provides a good result based on the predictions done on the images as strong positive or negative sentiments. The facial emotions detection is used for decision fusion mechanism. Based on this mechanism we can easily identify the sentiments based on the images.

III. CONCLUSION

In this literature survey, Sentiment analysis plays a pivot role in making decisions about the product. The analysis is made on the data collected from twitter. Location based Tweets are collected. Once collection of tweets is done, then the analysis process is carried on. Analysis is done not only on the text data but also on the emoticons as well as the images. Along with this the Tamil tweets are converted into English and can also be analyzed by using the Word Sense Disambiguation technique.

IV. REFERENCES

[1] Mohammad Samiul Islam, Faysal Ahmed, Rashedur M. Rahman, Syed Akib Anwar Hridoy and M. Tahmid Ekram, "Localized twitter opinion mining using sentiment analysis",Springer (2015).

[2] Justin Zhan and Xing Fang,"Sentiment analysis using product review data",Springer (2015).

[3] Carolina Bigonha, Mirella M. Moro, Marcos A. Gonçalves, Virgílio A. F. Almeida, Thiago N. C. Cardoso, "Sentiment-based influence detection on Twitter",The Brazilian Computer Society (2011).

[4] Borut Sluban, Igor Mozetič, Jasmina Smailović, Petra Kralj Novak, "Sentiment of Emojis",Plos one (2015).

[5] Hailin Jin and Jianchao Yang, Quanzeng You and Jiebo Luo, "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks",Association for the Advancement of Artificial Intelligence (2015).

[6] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau, "Sentiment Analysis of Twitter Data", NextGen Invent (NGI) Corporation(2012).

[7] Efthymios Kouloumpis, TheresaWilson and Johanna Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media(2011).

[8] James Spencer and Gulden Uchyigit, "Sentimentor: Sentiment Analysis of Twitter Data", School of Computing, Engineering and Mathematics, University of Brighton(2014).

[9] Aarati Patil and Srinivasa Narasimha Kini, "Location Based Sentiment Analysis of Products or Events over Social Media", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169, Volume: 4 Issue: 7,50 -55.

[10] Vasavi Gajarla and Aditi Gupta, "Emotion Detection and Sentiment Analysis of Images", Georgia Institute of Technology(2015).

[11] Kausikaa.N and V.Uma, "Sentiment Analysis of English and Tamil Tweets using Path Length Similarity based Word Sense Disambiguation", IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 3, Ver. I (May-Jun. 2016), PP 82-89.

[12] Gargi Mishra and Shivani Varshney, "Location Based Opinion Mining of Real Time Twitter Data", International Journal of Computer Science and Information Technologies, Vol.7 (4) , 2016, 1831-1835.

[13] Jianbo Yuan, Quanzeng You and Jiebo Luo, "Sentiment Analysis Using Social Multimedia", Springer(2015).