Scrutiny of Sentimental Analysis Techniques

Ikshita Dolly[1], Mandeep Singh[2]
[1]*Masters of Engineering Honours, Information Security, Computer Science Department, Chandigarh University*
[2]*Associate Professor of Computer Science Department, Chandigarh University*

Abstract - The sentiment analysis is the technique which is applied to analyze the sentiments of the input data. The sentiment analysis contains various steps like data pre-processing, feature extraction and classification. The classification defines the class in which the test data need to classify. In this review paper various techniques of sentiment analysis is reviewed and analyzed in terms of description and outcomes

Keywords - *Sentiment analysis, Classification, SVM, KNN*

I. INTRODUCTION

In order to mainly, determine whether a product is satisfactory for the users before it is delivered to them, the sentiment analysis (SA) has been developed. On the basis of the requirements of the users, the markets and industries develop their products and services. The factual data is processed, searched or analyzed with the help of textual information retrieval methods. The subjective properties of the components can be presented on the basis of various textual contents within the actualities [1]. The base of sentiment analysis (SA) includes opinions, attitudes, emotions, appraisals and so on. In order to develop new applications, various challenges have been faced while applying these techniques. The major reason due to which issues arise is the regular generation of huge types of data on various online platforms. The different types of positive or negative opinions are given by the users related to various objects which can help organizations in providing feedbacks that can be used in enhancing the quality of those objects. With the utilization of Natural Language Processing (NLP), various tweets, speech or text available on sources can be processed in sentiment analysis.

1.1. Challenges being faced in Sentiment Analysis

There are various challenges being faced when sentiment analysis is applied in various applications. Some of these challenges are [2]:

- This technique cannot differentiate amongst the fake or spam reviews that are present within the sources. There is no comparison made amongst the quality of reviews provided. The selection of outliers and identification of reputation of reviewer is also not provided here.
- Collaborative filtering which recognizes important key points and provides innovative ideas is limited within this technique.
- The automated content analysis and behavioral analysis are combined together to form a filter bubble. This helps in providing selection of related opinions and content in a very efficient manner. This is however a risk which is available in sentiment analysis methods and results in causing various issues [3].
- The various organizations and government can afford the presence of opinion mining software which is however not available to the other general people. This means that even when people have the democracy of producing and publicizing their content, they cannot have the access to analyze it.
- The behavior as well implicit data are integrated with opinion which helps in validating and providing analysis of the data that cannot be expressed in general form.
- Usability and user-friendliness of the tools are the permanent requirements which need to be maintained in these systems.

II. CLASSIFICATION

There is a need to assign a class of category to any kind of input that is given to us on the basis of the manner in which it is generated. The voices, faces, emails, grades and so on can be classified on the basis of various their characteristics. An automated process through which the units of texts can be divided and labeled amongst different categories which are called classes, is known as text classification method. The topic can be extracted from the text or can be involved within sentiment classification through this process [4]. The text available online can be classified as positive, negative or neutral with the help of NLP within sentiment analysis. The application which considers every individual word as a separate class and predicts the next work on the basis of previous content available is known as language modeling which is also a method of classification. The extraction and classification of useful features on the basis of their properties into discrete classes is the prior objective of classification. The utilization of hand-written rules is one method through which the text can be classified. Various applications are present that utilize hand-written rule-based classifiers as classification methods.

The manner in which the humans can learn from the experiences they had from their pasts is followed by the machine learning techniques. The knowledge is gathered and applied in order to make any kinds of decisions in future [5]. Within the applications of artificial intelligence and document classification, such various types of learning methods are utilized. There are mainly two steps that are followed which classification applied in machine learning. They are:

i. Utilizing the training dataset for learning the model.
ii. The trained model is applied to the test dataset.

There are various supervised classification methods amongst which any can be used within sentiment analysis method as it the part of text classification problem. Some of the most commonly used classifiers are explained below:

a. Naïve Bayes classifier: A classifier that comprises of Bayes theorem is known as Naïve-Bayes classifier. It is a simple probabilistic classifier in which it is assumed that the presence or absence of one feature in a document is not reliable on any other feature present within that document [6].

b. Support vector machines: In order to provide document classification, SVM has been proved to be the most efficient technique that can be used. The identifying of maximum margin hyperplane within the document which can separate the document vector in one class from that of another document which also has maximum margin is the main objective of SVM.

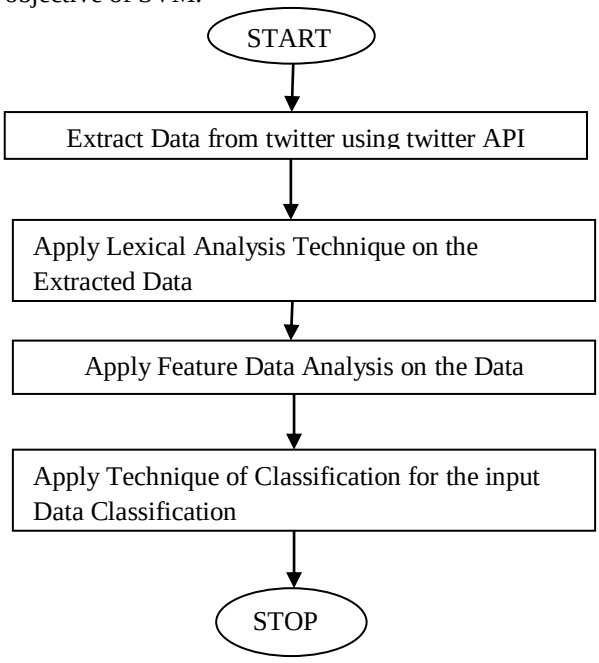

Fig 1: Proposed Technique

III. LITERATURE REVIEW

Wei Zhao, et.al (2017) presented in this paper [7], that there are numerous techniques that have been proposed related to opinion mining. However, there is a key challenge being faced in all of these techniques which relates to reviewing the orientation of sentence. Many frameworks that are used for the classification of the sentiments for product review are presented here. They are commonly represented as the weak supervision signals. There are two steps involved in the proposed framework amongst which, in step first, representation of high level learning is done where different rating information is confined for the distribution of general sentiment. In second step, one more layer that is classification layer is added to the embedding layer. This layer used labeled sentences for the supervision of fine tuning. The mode is reviewed through sentences that are of two kinds which are low level network structure such as convolutional feature extractors and long short-term memory. For the evaluation purpose the proposed framework is used where all the dataset is constructed on the basis of weekly labeled review that is 1.1M and also reviews are gathered from Amazon that is 11,754 reviews. It is concluded that the efficiency of the proposed framework is superior to existing frameworks.

Mondher Bouazizi, et.al (2017) presented in this paper [8], a novel technique to classify the texts that are gathered from Twitter in detailed manner. Amongst multiple sentiment classes, the text is classified through this technique. In order to assist the users for choosing appropriate features from various features provided, SENTA tool is introduced in this study. The identification of such features will help in making it easy to run the classification with the help of a readily available graphical user interface. In order to evaluate the experiments related to multi-class classification SENTA is utilized within this paper. On the basis of experimental results achieved it is seen that the multi-class classification is achieved with 60.2% of accuracy. Thus, by providing various comparisons it is seen that the performance of proposed technique is better than the existing approaches.

Ankit Kumar Soni, (2017) presented in this paper [9] the objective of developing a system which will help in extracting the useful information from raw data available in the application by using classification methods. This extracted data, which is mainly present in twitter micro blogging applications, has various sentiments present within it which need to be analyzed in order to analyze the views of users. There are innumerable tweets present which are daily posted on online applications which involve different opinions and sentiments of the users all across the globe. The languages in which tweets are posted are also different. The dataset which has Multilanguage tweets is not very easy to be handled by the

sentimental classification. There is no such technique proposed which can help in handling the multi-language data. In this paper, Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm. Amongst various algorithms, the results are compared which can help in analyzing the performance of various algorithms amongst each other and show which has provide to be better. It is seen through the results achieved that the proposed technique has provided better results in comparison to other existing approaches.

Aldo Hernández, et.al (2016) presented in this paper [10], a sentiment analysis technique that can help in predicting any kinds of future attacks that can possible arise within the web applications. From two sets of users, the tweets are gathered on daily basis which are used by this method. These users express their various issues in the form of content in the web applications. In order to predict the chances of attack, the daily gathered data is presented in statistic manner which can help in detecting the chance of any attack to occur. The sentiments of users and groups are analyzed collectively by the hackers here once the data is extracted by them from sources. Various experiments are conducted and it is seen that the proposed technique can help in providing an estimate of number of actual attacks and negative sentiments provided by the users in the form of tweets.

Venkata Sasank Pagolu, et.al (2016) proposed in this paper [11], the utilization of sentiment analysis and supervised machine learning principles in combined manner to analyze the sentiments of users on twitter platform. The data is extracted from twitter and the relation between the stock market movements of a company as well as the sentiment of the tweets are analyzed here. The users are motivate to invest in the stocks of a company when the tweets available on twitter are positive which might results in increasing the stock value of the organization in market. There is a direct relation seen amongst the tweets present by public and rise and fall of stock prices. The judgment of the types of sentiments present within the tweets posted by user is the prior object of the sentiment analyzer. On the basis of various experiential results achieved at the end, it is seen that the proposed technique provides better evaluation results in comparison to existing techniques.

Jaishree Ranganathan et.al (2017) proposed in this paper [12], a novel optimized and enhanced system through which the meta-actions can be generated. In order to generate these meta-actions, the Grabbing strategy (SARGS) algorithm based Specific Action Rule discovery is to be implemented. With the help of Twitter social networking data, all the meta-actions that generate algorithmic implementation within Apache Spark driven system and Hadoop driven system are reviewed in this paper. Corpus based sentiment analysis is performed on the social networking data here. The total time consumed by both the systems and their respective sub components in order to process the data is computed. As per the simulation results achieved it is seen that the computational time of proposed scheme is faster in comparison to other existing schemes.

Mitali Desai, et.al (2016) presented in this paper [13], that the highly unstructured data available on Twitter can be classified with the help of sentiment analysis. In order to perform sentiment analysis on Twitter data, the various techniques proposed are studied in this paper as well. On the basis of recognized parameters, for supervised machine learning techniques, a parametric comparison of the techniques is provided here. As per the experiments it is seen that there are domain and language based techniques provided for performing sentiment analysis of twitter data.

Rincy Jose, et.al (2015) proposed in this paper [14], a new mechanism in order to provide accurate sentiment classification of twitter messages which is based on lexical resources. In order to identify political sentiment for real time tweets, SentiWordNet lexical resource and Word Sense Disambiguation are applied. For achieving higher accuracy, a negation handling is utilized in this method in order to perform pre-processing step. For classification accuracy, there is around 1% of enhancement seen as per the results achieved from negation handling. There is around 2.6% of enhancement found within classification accuracy as per WSD results. The two novel products are compared in real time scenario as well by the twitter sentiment analyzer.

Huma Parveen, et.al (2016) presented in this paper [15], a study related to the extraction of sentiments from Twitter application in which the views and opinions are expressed through posts by users. A sentiment analysis on tweets is performed here through which some prediction related to business intelligence is provided. The movie data set present on twitter website includes content in the form of reviews, feedback and comments which needs to be analyzed. In different sections, the results of sentiment analysis performed on twitter data are presented such that the positive, negative and neutral sentiments are separated.

Monisha Kanakaraj et.al (2015) proposed in this paper [16], a Natural Language (NLP) based technique in order to improve the sentiment classification. To do so, the semantics within the feature vectors are added and in order to perform classification, the ensemble techniques are utilized. The accuracy of prediction can be enhanced by including semantically similar words and context-sense identities in the feature vectors. Simulation experiments are conducted and results are provided which show that the proposed technique outperforms other traditional techniques.

Author	Year	Description	Outcome	Future Scope
Wei Zhao, et.al	2017	A framework is proposed in order to evaluate all the dataset which is constructed on the basis of weekly labeled review that is 1.1M and also reviews are gathered from Amazon that is 11,754 reviews.	It is concluded that the efficiency of the proposed framework is superior to existing frameworks	The training dataset is weekly labeled which reduce accuracy
Mondher Bouazizi, et.al	2017	A novel technique to classify the texts that are gathered from Twitter in detailed manner. Amongst multiple sentiment classes, the text is classified through this technique.	On the basis of experimental results achieved it is seen that the multi-class classification is achieved with 60.2% of accuracy.	The classes in which data is classified are only two which can increased for batter classification
Ankit Kumar Soni	2017	Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm. Amongst various algorithms, the results are compared which can help in analyzing the performance of various algorithms amongst each other and show which has provide to be better.	It is seen through the results achieved that the proposed technique has provided better results in comparison to other existing approaches	To increase the accuracy of classification the CNN can be used to define classes
Aldo Hernández, et.al	2016	A sentiment analysis technique that can help in predicting any kinds of future attacks that can possible arise within the web applications. From two sets of users, the tweets are gathered on daily basis which are used by this method.	Various experiments are conducted and it is seen that the proposed technique can help in providing an estimate of number of actual attacks and negative sentiments provided by the users in the form of tweets.	The execution time is very high which can be reduced in future
Venkata Sasank Pagolu	2016	The data is extracted from twitter and the relation between the stock market movements of a company as well as the sentiment of the tweets are analyzed here.	On the basis of various experiential results achieved at the end, it is seen that the proposed technique provides better evaluation results in comparison to existing techniques.	The movie review dataset will be used for the classification
Jaishree Ranganathan et.al	2017	In this paper, a novel optimized and enhanced system is proposed through which the meta-actions can be generated.	As per the simulation results achieved it is seen that the computational time of proposed scheme is faster in comparison to other existing schemes.	The meta-actions are not generated accurately due to which improved is required in the technique
Mitali Desai, et.al	2016	In this paper it is studied that the highly unstructured data available on Twitter can be classified with the help of sentiment analysis for which numerous techniques are proposed which are studied here.	As per the experiments it is seen that there are domain and language based techniques provided for performing sentiment analysis of twitter data.	The performance of the algorithm can be tested on the structured data
Rincy Jose, et.al	2015	In this paper a new mechanism is proposed in order to provide accurate sentiment classification of twitter messages which is based on lexical resources.	There is around 2.6% of enhancement found within classification accuracy as per WSD results.	The priority technique can applied for the classification
Huma Parveen,	2016	In this paper, a study related to the	In different sections, the	The twippy application be

et.al		extraction of sentiments from Twitter application is presented in which the views and opinions are expressed through posts by users.	results of sentiment analysis performed on twitter data are presented such that the positive, negative and neutral sentiments are separated.	applied for the data extraction
Monisha Kanakaraj et.al	2015	In this paper, a Natural Language (NLP) based technique is proposed in order to improve the sentiment classification.	Simulation experiments are conducted and results are provided which show that the proposed technique outperforms other traditional techniques.	The classification technique can be applied to define classes.

Table 1: Table of Comparison

IV. CONCLUSION

In this review paper, it is concluded that sentiment analysis is the technique which is applied to analyze sentiments of the input data. The sentiment analysis process contains various stages like data pre-processing, feature extraction and classification. In the phase of pre-processing, the blank spacing or unwanted data is removed from the data. In the second step, the data features of the data are extracted by applying various algorithms like N-gram etc. In the last step, classification techniques will be applied which can classify data into certain number of classes. In future the technique will be designed for the sentiment analysis with the security parameters.

V. REFERENCES

[1]. Feng S., Wang D., Yu G., Yang C., Yang N., "Sentiment Clustering: A Novel Method to Explore in the Blogosphere" Advances in Data and Web Management, vol. 5446, pp. 332-344 , 2009

[2]. Andreevskaia A. , Bergler S. , "Mining WordNet For a Fuzzy Sentiment: Sentiment Tag Extraction From WordNet Glosses", Proceedings EACL-06, the 11rd Conference of the European Chapter of the Association for Computational Linguistics, pp. 209–216 ,2006

[3]. Fei, Geli, et al. "A Dictionary-Based Approach to Identifying Aspects Implied by Adjectives for Opinion Mining." 24th International Conference on Computational Linguistics. 2012.

[4]. M. Edkrantz, S. Truve, and A. Said Alan, "Predicting Vulnerability Exploits in the Wild," Cyber Security and Cloud Computing (CSCloud), Proc. 2015 IEEE 2nd International Conferen pp. 513-514, 2015

[5]. O. Owoputi, C. Dyer , K. Gimpel, N. Schneider, and N.A. Smith, "Improved part-of-speech tagging for online conversational text with word clusters" Proc. Conference of the North American Chapter of the Association for Computational Linguistics, 2013

[6]. K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, and N. A. Smith, "Part-of-speech tagging for Twitter: annotation, features, and experiments", Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers , vol. 2, pp. 42-47, 2011

[7]. Wei Zhao, Ziyu Guan, Long Chen, Xiaofei He, Deng Cai, Beidou Wang and Quan Wang, "Weakly-supervised Deep Embedding for Product Review Sentiment Analysis", 2017, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING

[8]. Mondher Bouazizi and Tomoaki Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", 2017, IEEE

[9]. Ankit Kumar Soni, "Multi-Lingual Sentiment Analysis of twitter data by using classification algorithms", 2017, IEEE

[10]. Aldo Hernández, Victor Sanchez, Gabriel Sánchez, Héctor Pérez, "Security Attack Prediction Based on User Sentiment Analysis of Twitter Data", 2016, IEEE

[11]. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", 2016, IEEE

[12]. Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva, "Action Rules for Sentiment Analysis on Twitter Data using Spark", 2017 IEEE International Conference on Data Mining Workshops

[13]. Mitali Desai, Mayuri A. Mehta, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey", International Conference on Computing, Communication and Automation (ICCCA2016)

[14]. Rincy Jose, Varghese S Chooralil, "Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Word Sense Disambiguation", 2015 International Conference on Control, Communication & Computing India (ICCC)

[15]. Huma Parveen, Prof. Shikha Pandey, "Sentiment Analysis on Twitter Data-set using Naïve Bayes Algorithm", 2016, IEEE

Monisha Kanakaraj and Ram Mohana Reddy Guddeti, "NLP Based Sentiment Analysis on Twitter Data Using Ensemble Classifiers", 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)