Novel Approach for Sentiment Classification by Ontology with Swarm Intelligence Optimization: A Review

Aditya Bhardwaj[1], Uma Tyagi[2]
[1]*M.Tech, Dept. of Computer Science And Engineering, Meerut Institute of Engineering And Technology, Meerut, India*
[2]*Asst. Prof., Dept. of Computer Science And Engineering, Meerut Institute of Engineering And Technology, Meerut, India*

Abstract- Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. However, they are now all under the umbrella of sentiment analysis or opinion mining. While in industry, the term sentiment analysis is more commonly used, but in academia both sentiment analysis and opinion mining are frequently employed.

Keyword- Sentiment, Optimization, Learning, Analysis.

I. INTRODUCTION

Technology has put its feet in every field whether it is industrial sector, manufacturing, media, healthcare, communications or social media. Trillions of data is coming from every sector at an abrupt rate. So to handle this large volume of data having huge variety and coming at a single instance is very difficult and it is termed as Big Data. Interrogation include different fields like acquisition, analysis, repository search,transfer rate of data, sharing, perception, renewing and querying the data with confidentiality [1].Basic purpose of Big Data is predictive analysis, user behaviour analysis, or use different other progressive data analytic methods in which query data is derived from a sample of data set. Scrutiny of sample data can explore new interrelationships, helpful for realizing market trends, warfare crime, prohibit diseases.

A. Sentiment Analysis

Sentiment analysis is that process in which sentiments are identified from a text unit through some technique i.e. natural language processing, statistics or machine learning [3]. It can be used in different fields like politics, sociology, psychology, sports, brands, entertainment because tweets include mass opinion. For eg. In politics through sentiment analysis we can analyze the trends, can know views about party policies. It helps in predicting polarity because tweets include the public opinion. Similarly it can be used by companies to know the market trends and public poll for products. In simple terms, predicting whether the given sentiment or review is positive or negative at a huge rate is known as sentiment analysis.

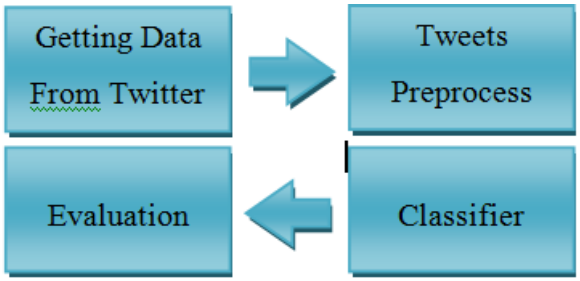

Fig.1: Sentiment Analysis Process.

B. Need of Sentiment Analysis

Social media sentiment analysis can be an excellent source of information and can provide insights that can:

- It determines the marketing strategy
- It improves the campaign success
- It improves the product messaging
- It Improves the customer services
- Test the business KPIs
- It is used to Generate the leads in the business

C. Types of Sentiment Analysis

1. Manual Processing

Human interpretation of sentiment is definitely the most mature and accurate judge of sentiment. However, it still isn't 100% accurate. Very few vendors still use this process without the additional use of a tool. This is due to the prolific growth of social media.

2. Keyword processing

Keyword processing algorithms assign a degree of positivity or negativity to an individual word, then it gives and overall

percentage score to the post. For example, positive words, great, like, love or negative words: terrible, dislike.

The advantages of this method are that it is very fast, predictable and cheap to implement and run.

However, there are numerous disadvantages including dealing with double negatives or positives or different meanings of words, for example, the use of a word such as 'sick' (to mean either "ill" or to mean "awesome"). Not to mention, different researchers may assign difference percentages of positive or negative to a word. More often the issue is that it does not deal with multiple word/context issues or non-adjective words.

Most vendors represented in Australia use a keyword processing algorithm.

3. Natural Language Processing

NLP also called: text analytics, data mining, computational linguistics) NLP refers to computer systems that process human language in terms of its meaning. NLP understands that several words make a phrase, several phrases make a sentence and, ultimately, sentences convey ideas. NLP works by analysing language for it's meaning.

II. RELATED STUDY

Ref. no.	**Year of publication**	**Summary**	**Algorithm used**	**Pros and Cons mentioned**	**Future Scope**
[1]	2016	This paper represents the senticircles-lexicon based approach for the analysis of the sentiments of the content on twitter. It considered the co-occurrence pattern of the word in different context from twitter to catch their semantics and refresh their strength and polarity. This hypothesis permits to recognize the sentiments at two level 1) Entity-level and 2) Tweet-level.	Senticircle a lexicon based approach.	This method gives average result in analyzation process.	Their approach is to perform better than the state-of-the-art. 4-5 % in accuracy in two database but fall behind 1 % in third data sets.
[2]	2014	This paper proposes a method mine for twitter data. Here we use data mining algorithm for determination process. This technique is used to determine the price of the selected companies listed in the 30 companies. NASDAQ and the new York stock exchange can actually be estimated by the 15 million records of twitter message. Extracting twitter message data through NLP will help in the process of estimation NLP is used to discover pattern between public sentiment and real stock price.	Data Algorithm.	NLP method gives results on real stock price.	To increase the accuracy percentage from prediction to actual investment income.
[3]	2014	This hypothesis gives an examination on the sentiment analysis for the client which analyzes the information in the form of the number of tweets where opinion is unstructured. Here we first pre prepared that information which have feature vector. Then we select that feature and connected machine learning	Machine learning with SVM.	Showing issues in the analyses of same meaning words.	In future performance can be improved by using any other method.

		based on classification. After that synonyms are extracted by SVM along with Semantic Orientation. Then finally we measure the performance as far as precision and exactness.			
[4]	2014	This theory introduced a Senticircle, a novel lexicon-based approach. This will consider the contextual and conceptual semantics of the words while analysis. Here we evaluate three twitter dataset with three different sentiments together and check the results. It has seen clear that our approach is better to applied in the analyzation of the sentiments.	Senticircle approach.	Senticircle is only giving average results when we precedethe analyzation of sentiments from different contents.	To compare approach with other approach which take semantic into account for sentiments detection like SenticNet.
[5]	2013	For the efficient sentiment analysis, this hypothesis proposes a technique based on deployment of original ontology to the post of twitter. Post are not simplify portrayed by a sentiment score however instead get a sentiment review for every notion in the post in machine learning. This is novelty of the reason. This proposition brings about the point by point analysis of post opinions regarding a specific theme.	Ontology Approach.	It limited only for textual sentiment analysis.	The integration of a custom-built sentiment classifier that will substitute Open Dover in our architecture.
[6]	2013	In this paper we use to analyze the twitter posts with the assistance of electronics devices like versatile, tablet and so on by using machine learning approach. This approaches the machine learning technique for the analysis of sentiments. These arrangements with the identifying and classifying or sentiment communicated in the given content.	Machine learning.	There are some issues regarding identification of emotional keywords. Shows difficult to handle misspelling and slang words.	Developing New feature vector to extract people's opinion about product.
[7]	2013	This theory introduced n-grams to reduce supervised features and statically analysis to develop twitter-specific lexicons to the analysis of the sentiments. This reduced Twitter-specific lexicon augmented with brand-specific terms for Brand related tweets. It shows that lexicon sets reduces modeling complexity , and maintain a high degree of coverage over twitter and improves the analysis .	SVM and DAN2 .	Reduces problem complexity Model size reduction.	Experimentation on additional brand and similar Twitter corpuses. This approach allows brands to monitor

					sentiment in twitter.
[8]	2013	In this paper a frame work is intended for the examination of the twitter assessments. This framework utilizes expound bootstrapping outfit to subdue class awkwardness, sparsely, and representational to investigate the opinions of the issue. A content investigation framework is proposed for Twitter assessment examination. Because of substantial he class lopsidedness in a multi-class issue, conclusion order stays subtle. These issues are hazardous since many types of online networking. While doing tests, result demonstrates that framework approach is more precise and adjusted in the prediction crosswise over opinion classes as contrast with different devices. Bootstrapping framework helps to build sentiment time series that are better able to reflect events like positive or negative sentiments.	Bootstrapping are used to analyze the sentiments.	Framework used in it provides better result in identifying positive and negative sentiment.	Extending the expansion parameters and to improvement on searchmethod.Expounding the BPEF by giving extensible nature.
[9]	2013	This paper introduces a novel i.e. Aspect-based sentiment analysis. The main focus of this novel is Short text mainly focused on Twitter post or messages. Here we use different algorithms for the analyses of polarity detection and sentiments. This novel shows that it is advantageous for unigram state-of-the-art baseline. This novel has high performance with useful functionalities and features.	Aspect-based sentiment analysis.	Results shows that this system gives high performance.	Building up a multi-domain context dependent lexicon and finding more features for the semantic analysis.
[10]	2012	In this paper, there is an introduction semantic feature into the training sets for sentiment analysis. Addition of semantic concept as additional features for each removed entity. This procedure helps in measuring the correlation of the concept with negative\positive sentiments. We use the approach to expect sentiments for three informational index of twitter.	Addition of semantic concept as additional feature.	Measures relation between negative /positive sentiments properly and with convenient way.	Enlargement of the accuracy rate for both sentiments and unigrams lines.
[11]	2012	This paper gives an approach where Preprocessing and characterization in view of their passionate substance as	Based on classification and pre-processing.	Less accuracy**.** Sleekness in the datasets.	International expressions and foreign

		positive/negative and unessential investigation. The exhibitions depend on the precision and exactness .every one of the opinions is investigated with characterization and pre-processing.			words classificatio n in more details.
[12]	2012	This paper proposed an automatic training based sentiment analysis on tweets contain emoticons or sentiment based word. There are some sets introduced used to evaluate the tweet left by automatic classifier. Naive Bayes algorithm is used to classify the enabled tweets. Emoticon based, word base and hybrid method with different criteria is used for the automatic classification. This technique gives 90 % accuracy. But the combination of the technique gives improved accuracy.	Emoticon based technique , Word based technique and hybrid method.	In this paper hybrid technique is used for sentiment analysis and gives better results.	Accuracy can be enhanced by using other method.
[13]	2012	This paper gives an application on Arabic sentiment analyses. This is done by putting a sentiment classification for Arabic twitter messages. The messages are analyzed to provide there sentiments weather positive or negative. All the data are collected from the social site Twitter. It has its own importance in the region of middle east because on that region only Arabic language is used.	Introduced an Arabic application or analyzing sentiments in Arabic language.	Limiting research is present in this analysis**.**	Creation of hybrid approach which was the combination of ML and SO. This will happen by adding more comprehensi ve list (all negative, positive sentiments) In Egyptian language.
[14]	2011	This hypothesis represents the three techniques to investigate high volume information. The techniques are 1) SA based on subjects to removes maps and measures opinions of the clients. 2) Stream analysis that identifiesinteresting tweets based on their density, negativity, influence and attributes. 3) Pixel cell-based sentiment calendars and high density geo maps that visualize expansive volumes of information. This technique is utilized to the assortment of twitter information.	Technique used are as, SA on topic, stream analyses and pixel cell-based sentiments.	This is a better technique explored for high volume data. This can be used for various varieties of data.	To incorporate information about opinion association to findfeatures and to visualize them.

[15]	2010	This hypothesis represents the machine learning approach to the classification of sentiments of twitter message. In this hypothesis each tweet will classified into two classifications say polar and non-polar. Polar sentiments are those which are with positive or negative sentiments, left as non-polar. They represent text normalization strategy for noisy tweets and classified them concerning the polarity.	Machine learning approach is used to analyze the sentiments.	This technique gives better result as compare to the n- gram features.	To learn lexicon extracted from tweets generative model has to build up. This will help to represent the similar concepts like love,lv, loveee and luv. An important direction for future is depending upon the entity-based concept model with lexicon-semantic knowledge.

III. CONCLUSION

Sentiment Analysis (SA), also called Opinion Mining, is currently one of the most studied research fields. It aims to analyze people's sentiments, opinions, attitudes, emotions, etc., towards elements such as topics, products, individuals, organizations, services, etc. Different techniques and software tools are being developed to carry out Sentiment Analysis.

The goal of this work is to review and compare some free access web services, analyzing their capabilities to classify and score different pieces of text with respect to the sentiments contained therein. For that purpose, three well-known collections have been used to perform several experiments whose results are shown and commented upon, leading to some interesting conclusions about the capabilities of each analyzed tool.

IV. REFERENCES

[1]. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. Information Processing & Management, 52(1), 5-19.

[2]. Bing, Li, Keith CC Chan, and Carol Ou. "Public sentiment analysis in Twitter data for prediction of a company's stock price movements." e-Business Engineering (ICEBE), 2014 IEEE 11th International Conference on. IEEE, 2014.

[3]. Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014 seventh international conference on. IEEE, 2014.

[4]. [Saif, Hassan, et al. "Senticircles for contextual and conceptual semantic sentiment analysis of twitter." European Semantic Web Conference. Springer, Cham, 2014.

[5]. Kontopoulos, Efstratios, et al. "Ontology-based sentiment analysis of twitter posts." Expert systems with applications40.10 (2013): 4065-4074.

[6]. Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Computing, Communications andNetworking Technologies (ICCCNT), 2013 Fourth International Conference 2013.

[7]. Ghiassi, Manoochehr, James Skinner, and David Zimbra. "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network." Expert Systems with applications 40.16 (2013): 6266-6282.

[8]. Hassan, Ammar, Ahmed Abbasi, and Daniel Zeng. "Twitter sentiment analysis: A bootstrap ensemble framework." Social Computing (SocialCom), 2013 International Conference on. IEEE, 2013.

[9]. Neethu, M. S., and R. Rajasree. "Sentiment analysis in twitter using machine learning techniques." Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference 2013.

[10]. Saif, Hassan, Yulan He, and Harith Alani. "Semantic sentiment analysis of twitter." The Semantic Web–ISWC 2012 (2012): 508-524.

[11]. Gokulakrishnan, Balakrishnan, et al. "Opinion mining and sentiment analysis on a twitter data stream." Advances in ICT for emerging regions (ICTer), 2012 International Conference on. IEEE, 2012.

[12]. Lima, Ana Carolina ES, and Leandro Nunes de Castro. "Automatic sentiment analysis of Twitter messages." Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference on. IEEE, 2012.

[13]. Shoukry, Amira, and Ahmed Rafea. "Sentence-level Arabic sentiment analysis." Collaboration Technologies and Systems (CTS), 2012 International Conference on. IEEE, 2012.

[14]. Hao, Ming, et al. "Visual sentiment analysis on twitter data streams." Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. IEEE, 2011.

[15]. Celikyilmaz, Asli, Dilek Hakkani-Tür, and Junlan Feng. "Probabilistic model-based sentiment analysis of twitter messages." Spoken Language Technology Workshop (SLT), 2010 IEEE. IEEE, 2010.