Intelligent Opinion Polarity and Analysis in Discovering Product Related Customer Reviews: A Survey

Karuna Sahay [1*], Kaptaan Singh [2], Amit Saxena [3]

[1,2,3]Department of CSE, Truba Institute of Engineering & Technology, Bhopal, M.P., India

Corresponding Author: karuna.sahay28@gmail.com, Mobile: +918527987435

*Abstract--*Due to rapid advancements in Social media consumer interactions are increasing at faster rate. Twitter has now a days become a social media platform for industries, individuals, educational institutes and organizations who have a strong educational, political, industrial, social, banking or economic concern in maintaining and enhancing their social status and reputation. Posts are generally composed of poorly structured, incomplete, and noisy sentences, irregular expressions, non-dictionary terms, and ill-formed words. The problem is some customers given rating contrast with their comments. The other reviewers must read many comments and comprehend the comments that are different from the rating. Opinion Mining is the computational detailed investigation of people's attitudes, opinions, and emotions concerning of issues, events, topics or individuals. This paper represents the survey of customer feelings related to online product with their opinion polarity and analysis.

Keywords: Sentiment analysis, Opinion mining, Machine learning, Social Media, Support Vector Machine, Sentiment Polarity

I. INTRODUCTION

Twitter, with more than 400 million monthly active users and over 550 million tweets per day is most widely used social media applications. The main aims of opinion mining classification is to recognize the opinion polarity of a consumer tweet as neutral negative, or positive for products. The opinion entity can be product reviews, any educational events, individual's comments, or historical topics. These opinion are likely to be hidden from analyses. Opinion Mining[1] is an unending field of research in document mining field. Opinion mining will review different post of users and mine their opinion about related subjects. Opinion mining classification aims to categorize the opinion polarity of a tweet as negative, positive, or neutral. Posts are generally composed of poorly structured, incomplete, and noisy sentences, irregular expressions, non-dictionary terms, and ill-formed words. Preprocessing means removing URLs, removing stop words, and replacing negations from users post. A series of pre-processing are applied to reduce the amount of noise in the posts before feature selection. Pre-processing is accomplished comprehensively in existing methodologies, specifically in machine learning-based methods.

Opinion Mining (OM) also termed as Sentiment Analysis (SA) is the computational analysis of public's attitudes, opinions, and sentiments or opinions concerning an entity.

The object can signify events, individuals, or topics. These subjects are furthermost likely to be concealed by analyses. The two expressions OM or SA are interchangeable and express a common meaning. Opinion Mining is an unending field of research in document mining field. Opinion mining will review different post of users and mine their opinion about related subjects. The Clustering and natural language processing procedure will be applied for opinion mining. A portion of opinion mining denotes using of natural language processing (NLP)[2] by suggested dissimilar method of dictionary for sentimentality analysis of text data as lexicon, corpus, and specific language dictionary.

The rest of the paper is organized as follows.
Section 2 represents the problem identification. Section 3 provides literature survey. Section 4 provides conclusion of the paper.

II. PROBLEM IDENTIFICATION

The problem is some customers given rating contrast with their comments. The other reviewers must read many comments and comprehensive the comments that are different from the rating. The challenge is in process of opinion mining or sentiment analysis that is unstructured and noisy data on website. The datasets used in OM are an essential problem in opinion review

field. The foremost issues that interfere with the user comfort and security are confidentiality breach, groups lacking opt-in options, disorder created out of various groups in which a consumer is a member of and struggle in managing group ideologies. The purpose of utmost of the groups in widely held social networking societies such as Twitter is misleading by uproars like spams and advertisements, and other clatters that obstruct with a group members interest.

The challenging task related with the groups is the administration of group guidelines. The furthermost widespread social networking site Twitter has group's members with over 100K. Consequently, it turn out to be difficult for the supervision to track the followers violating group guidelines. This represents that there occurs a need for a measure to classify the member posts based on suitability and group behavior. A scheme to control the entry of inappropriate messages within a group is very much necessary for the smooth operational of a social networking sites. These days, a company or organization make available a business service which essentials to get opinion from customer. The customer assessment is essential to increase better service for company.

III. LITERATURE SURVEY

Anh Vo et. el.[3] Suggests a technique to summarize and extract opinion and aspect of product form post of consumer in large dataset. The experimental outcome represents that the proposed method achieves F1-scores of 0.774 for laptop product review and 0.714 for camera product reviews. From results the author examined the usefulness and precision of proposed method. Consumer opinion analysis on online sold product is not an easy task, due to numerous assessments on Internet. Consumer opinion analysis problem can be addressed by applying syntax of reviews and exploring the appearance. It can also be solved by looking for expressed feeling in the post.

Opinion mining technique called Aspect-based evaluation is most widely used method. Most of the research are conducting in product aspect extraction from posts. The problem of opinion mining are complex algorithm and lack of handling large data set.

The datasets used in OM are an essential problem in opinion review field. The foremost main sources of data are from the consumer product reviews. The people's reviews are significant to the industry holders as they are takings business decisions agreeing to the analysis outcomes of users' thoughts about their products. The assessments sources are primarily review sites. OM can similarly be applicable on news articles, political debates or stock markets. For example in political debates, the investigators could figure out public's opinions on a certain political

parties or election candidates. The election outcomes can also be forecast from user's political posts. The micro-blogging sites and social network sites are considered a good source of customer information because people discuss and share their opinions nearly a certain topic freely. They are similarly used as data sources in the OM process.

At the present time, an establishment or association make available a commercial service which essentials to acquire advice from consumer. The consumer review is significant to progress service for organization or establishment, which have together nearby opinion and undefended opinion. The open estimation means the observation as manuscript which appearances emotion and statement in a straight line from consumer. Nevertheless, the company has various innards or group to calculation themselves by ranking and overall assessment for a category of service area which nearby are several consumer who requirements to analysis.

With the speedy extension of corporation or association have additional services and goods online and increase consumer contentment. The supplier will read consumer assessment and other consumers who necessity to procedure amenities or goods will read assessment to express opinions on the facilities. The numeral of consumer assessment is accumulative or vast from website, ha blogs, opportunities and community broadcasting, which the amenities or invention is motivating. Therefore, countless consumers will read comment randomly which is tough to read altogether comments and create choice the facilities or goods. If consumer reads an insufficient reviews, consumer might acquire opinion assessment to be partiality. Therefore, opinion excavating is a procedure of field area of info abstraction from text dispensation, which is assistance and several occasions to progress or change factor to business grind by this investigation. Investigation of judgement has equal of sentimentality from sentiment word and considered total of correspondence or collection with the benevolent of expression as optimistic or undesirable called sentiment polarity[4].

In numerous solicitations, it is imperative to contemplate the situation of the manuscript and the consumer partialities. Using TL techniques, we can use associated data to the province in inquiry as a training information. The natural language processing apparatuses can be applied to simplify the OM progression. It provides improved natural language thoughtful and thus be able to benefit harvest additional correct outcomes of OM. That is why OM need to create more research on context-based OM. Using NLP software to strengthen the OM process has fascinated academicians recently and still necessities of some enhancements.

Sentiment Analysis classification
The sentiment analysis classification pictorial representation is shown in figure.

Analysis Techniques
The following analysis techniques are used for sentiment analysis classification.
1. Lexicon-based technique
2. Machine learning technique

Machine Learning Technique:
This is unique and most suitable methods presented in sentiment analysis to categorized document or sentence into negative, positive or neutral groupings.

A. Supervised Machine Learning Approach
This approach is used for classifying the sentence or document into negative, positive and neutral classification. Maximum Entropy (ME), Naïve Bayes (NB), Support Vector Machine (SVM) are the furthermost useful systems. Supervised learning can be additional classified into Rule based classifier, Probabilistic classifier, Decision tree classifier, and Linear classifier

(a) *Probabilistic Classifier*: The different methods are used in Probabilistic Classifier.
(i) *Naïve Bayes Classifier (NBC)*: The Naive Bayes classifier is the simplest and, universally applied classifier. This classification method is used in numerous real world difficulties like email Auto Grouping, email sorting by priority, Sentiment analysis, email Spam Detection, Document Categorization and Sexually explicit content detection Naïve Bayes Classifier is used. The most important advantage of Naïve Bayes is it needs little processing memory and a smaller amount of time for implementation. A naive Bayes classifier is a modest probabilistic classifier constructed on Bayes' theorem and is principally matched when the dimensionality of the inputs are great [17].

(ii) *Bayesian Network (BN)*: Assumption about the NB classifier is that it is independence of the features. The other extreme assumption is to assume that all the features are fully dependent. This leads to the Bayesian Network model which is a directed acyclic graph whose nodes represent random variables, and edges represent conditional dependencies. BN is considered a complete model for the variables and their relationships. It is not used frequently in Text mining, because the computation complexity of BN is very expensive. [5]

(iii) *Maximum Entropy Classifier (ME)*: In numerous Natural Language processing uses Maximum Entropy (ME) classifier demonstrated in effect and use. Nothing like Naïve Bayes Classifier the Max Entropy ensures not take up that the features are tentatively independent of each other. Max Entropy classifier can be applied for opinion analysis also to resolve a great variety of text classification difficulties such as topic classification and language detection, etc. Additional time is necessary to train as equate to Naïve Bayes.

Advantages of using Max Entropy are performance, consistency, efficiency, accuracy, consistency and flexibility

(b) *Linear classifier*: Linear Classifier methods are represented in following section.
(i) *Support Vector Machine*: Support vector machine be able to be used for regression and classification purpose.
Commonly in text categorization SVM with machine learning classifier is applied. It has been observed that Support Vector Machines (SVMs) is extremely effective at traditional text cataloging. SVM take large-margin, relatively probabilistic classifiers. Foremost principle of SVM is to define linear separators in examine space which can finest discrete the dissimilar classes. The Tweets to be categorized are transformed into word vectors. SVM seek out a decision surface to dispersed training data point addicted to 2 classes and mark decision constructed on support vectors [6].

(ii) *Neural Network*: It consists of many neurons where the neuron is its basic unit. The inputs to the neurons be there represented by the vector over line Xj which is the word occurrences in the jth document. There exist a group of weights A which are connected with each neuron applied in direction to calculate a gathering of its inputs. Constructed on weights inputs and output is produced. [7]

(c) *Decision Tree classifier*: Decision tree constructs regression or classification prototypes in the arrangement of a tree structure means that the Decision tree classifier make available a hierarchical decomposition of the training posts data space in which a circumstance on the attribute significance is used to applied the posts data.

It breaks down a posts dataset into minor and smaller subcategories while at the same time connected decision tree is incrementally established. The partition of the data space is completed recursively up until the leaf nodes comprehend convinced minimum numbers of records which are applied for the determination of classification. [8]. And as a final point get a tree with decision methods and leaf nodes.

(d) *Rule based classifier*: Rule constructed classifier is very simple the individual rule applied here is concerned with excitements in the text. This scheme returns the negative and positive emoticons coordinated in text like positive emoticons are apprehensive with positive ness in text although negative emotions are apprehensive with negativeness. [9]

In rule constructed classifiers, the data space is molded with a group of guidelines. The left hand side characterizes a condition on the feature set conveyed in disjunctive normal form while the right hand side is the main class label. The circumstances are on the term occurrence. Term absence is rarely used because it is not explanatory in sparse data. [10]

B. Unsupervised Machine Learning Approach

To catalogue the document in classifications text classification is useful. In text classification, it is informal to accumulate unlabeled documents but occasionally it is problematic to create labeled documents, this struggle can be circumvented by using the unsupervised technique. It divisions the document into sentence and catalogue, each sentence using keyword list of each category's. As sentiment tokenized words and phrases possibly used for sentiment arrangement in an unsupervised method so it is not difficult to conceive that. Unsupervised machine learning methods don't use training data group for arrangement. Resources to categorize data into groupings

Collecting procedures like K-means assembling, Hierarchical clustering are used. The further unsupervised method Semantic Positioning also use to produce perfect result for grouping.

The Lexicon-based Approach

To compute the sentiment polarity for an examination using the semantic orientation of words or sentences in the examination the lexicon-based methodology is used**.** The subjectivity and opinion in tokenizer text is measure in semantic orientation. It be contingent on conclusion the opinion lexicon which is applied to analyze the text.
There are two approaches in this methodology.
(a) *Dictionary based approach*: This method be contingent on finding opinion seed words, and then examinations the dictionary of their antonyms and synonyms.
(b) *Corpus based approach*: Activates with a seed list of opinion words, and then discoveries other opinion words in an outsized corpus to help in outcome opinion words with environment unambiguous orientations.
(a) Statistical approach
(b) Semantic approach

Text View

Sentiment analysis has been inspected primarily at three levels- sentence level, document level, feature level.

Document Level Sentiment Analysis

This level of sentiment arrangement[11], a particular review about a particular topic is deliberated. Document level analysis is not necessary in blogs and forums. The imperative thing in this type of analysis is objectivity or subjectivity arrangement. The document level sentiment arrangement has its own disadvantages and advantages. The improvement of document level sentiment analysis is that get an overall polarity[12] of opinion tokenized text about an individual entity from a document and disadvantage is that the dissimilar emotions about dissimilar features of an object could not be take out separately. [13] In document level classification both unsupervised and supervised learning approaches can be used. Somewhat supervised learning procedure like Naive Bayesian[14], support Vector Machine,

can be applied to train the classification. For testing data training and, the reviewer rating (in the form of 1-5 stars), can be applied. The features that can be used for the machine learning are duration frequency, adjectives from part of speech cataloging, opinion words and phrases, syntactic dependencies, negations, etc. The unsupervised learning can be completed by removing the opinion words private a document. The point-wise mutual data can be used to find the semantics of the removed words.

Sentence Level Sentiment Analysis

In this sentence level arrangement the identical document level organization approaches can be used for the same difficult. The polarity for each sentence is considered in the sentence level sentiment analysis. The problem is to find subjective and objective sentences. The subjective sentences comprise[15] opinion words which benefit in formative the sentiment about the object. Afterward the polarity classification is completed into negative and positive classes. In this case of modest sentences, an only sentence accepts a single opinion about an object. Sentence level sentiment classification is not necessary in composite sentences. Significant that a sentence is negative or positive is of slighter use than significant the polarity of a specific feature of a product. The benefit of sentence level examination lies in the sub objectivity organization. The traditional procedures can be used for the training processes.

Feature Level Sentiment Analysis

This methodology is the best and important methodology to opinion mining. The phrase level arrangement is done by establish out the phrases that comprise opinion words. This investigation has disadvantage and advantage also. In more or less cases, the particular opinion about an object can be appropriately removed. But then again in some new cases, where circumstantial polarity similarly problems, the outcome may not be completely accurate. Negation[16] of words be able to take place locally. In such cases, this level of opinion analysis suffices. Nevertheless if there are sentences with opposing words which are distant apart from the opinion words, this phrase level analysis is not necessary. Likewise long range dependences are not considered in this method. The words that look like very near to for each other are well-thought-out to be in a phrase.

IV. CONCLUSIONS

The natural language processing implementations can be applied to facilitate the OM process. With the growing inspiration of online sentiment analysis and reviews on customers, the competence to detect dishonest online appraisals is crucial. It provides improved natural language understanding then can help produce further improved accurate outcomes of OM. In numerous applications, it is significant to consider the context of the text data and the user preferences. The paper provides the survey of different

opinion analysis and polarity related to customer review related to online product.

REFERENCES

[1] Wararat Songpan, The Analysis and Prediction of Customer Review Rating Using Opinion Mining, IEEE SERA 2017, pp. 71-77

[2] Arno Scharl, David Herring, Walter Rafelsberger, Alexander Hubmann-Haidvogel, Ruslan Kamolov, Daniel Fischl, Michael Föls, and Albert Weichselbraun, "Semantic Systems and Visual Tools to Support Environmental Communication", IEEE SYSTEMS JOURNAL, VOL. 11, NO. 2, JUNE 2017, pp. 762-772

[3] ANH-DUNG VO , QUANG-PHUOC NGUYEN , AND CHEOL-YOUNG OCK, Opinion-Aspect Relations in Cognizing Customer Feelings via Reviews, IEEE 2018, Vol-6, pp.5414-5426

[4] Kamps, J., Marx, M., Mokken, R. J.*Using WordNet to Measure Semantic Orientation of Adjectives*. LREC 2004. Volume IV, pp. 1115-1118.

[5] Andreevskaia, A., Bergler, S., Urseanu, M.*All Blogs Are Not Made Equal: Exploring Genre Di_erences in Sentiment Tagging of Blogs*. International Conference on Weblogs and Social Media (ICWSM-2007), Boulder, CO. 2007.

[6] Vandana V. Chaudhari*****, Chitra A. Dhawale****** and Sanjay Misra," Sentiment Analysis Classification: A Brief Review", I J C T A, 9(23) 2016, pp. 447-454

[7]ANH-DUNG VO , QUANG-PHUOC NGUYEN , AND CHEOL-YOUNG OCK, "Opinion_Aspect Relations in Cognizing Customer Feelings via Reviews", IEEE 2017, pp. 5415-5427

[8]ATHIRA U, AND SABU M. THAMPI, "Linguistic Feature Based Filtering Mechanism for Recommending Posts in a Social Networking Group", IEEE 2018, pp. 4469-4484

[9] S. 1. Wu, R.D. Chiang and Z.H. Ji, Development of a Chinese opinion mining system for application to Internet online forum, The Journal of Supercomputing, Springer US[Online], 2016.

[10] Z. Li, L.Liu and C.Li, Analysis of customer satisfaction from Chinese reviews using opinion mining, Proceeding of the 6th IEEE International Conference on Software Engineering and Service Science(ICSESS). 2015, pp.95-99.

[11] Q.Su, X.Xu, H.Guo, Z.Guo, X. Wu, X. Zhang and B.Swen. Hidden Sentiment association in Chinese web opinion mining. Proceeding of the 17th International Conference on World Wide Web, 2008, pp.959-968.

[12] R.M. Duwairi and I. Qarqaz, Arabic Sentiment Analysis using Supervised Classification. Proceeding of 2014 International Conference on Future Internet of Things and Cloud. 2014, pp. 579-583.

[13] H.S. Le, T.V. Le and T.V. Pham, Aspect Analysis for Opinion Mining of Vietnamese Text. Proceeding of International Conference on Advance Computing and Application, 2015, pp.118-123.

[14] V.B. Raut and D.D. Londhe, "Survey on opinion mining and summarization of user review on web", International Journal of Computer Science and Information Technology, Vol. 5(2), 2014, pp. 1026-1030.

[15] Fiaidhi, O. Mohammed, S. Mohammed, S. Fong, and T.H, Kim, Opinion Mining over twiiterspace: Classifying tweets programmatically using the R approach. Proceeding of the 7th International Conference on Digital Information Management, 2012, pp. 313-319.

[16] L. Lin, 1. Li, R. Zhang, W. Yu and C. Sun, Opinion mInIng and sentiment analysis in social networks: A retweeting structure-aware approach. Proceeding of the 7th International Confernece on Utility and Cloud Computing, 2014, pp.890-895.