JCSE International Journal of Computer Sciences and Engineering Open Access

Review Paper Vol.-7, Issue-7, July 2019 E-ISSN: 2347-2693

A New Approach to Handling Erroneous Reviews in Opinion Mining

Kirti Kushwah[1*], Rajendra Kumar Gupta[2]

[1,2]Department of Computer Science & Engineering, Madhav Institute Of Technology & Science, 474005, India

Corresponding Author: kirtikushwah.cse@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i7.5462 | Available online at: www.ijcseonline.org

Accepted: 12/Jul/2019, Published: 31/Jul/2019

Abstract: In the areas of marketing and electronic advertising, Opinion Mining has a broader domain. The advertiser must analyze the performance / popularity of the advertisements he has posted on the site. The mechanism based on the star rating can be fraudulent, due to robots or automatic responders. Therefore, it is necessary to analyze the current entity system or products using reviews (comments). Opinion Mining refers to the extraction of those lines or phrases in the huge raw data that express an opinion. On the other hand Sentimental Analysis is the analysis of feelings identifies the polarity (sentiment) of the opinion that is extracted from the review. Today, social networking sites and online shopping sites are used by users to express their opinion on products, events, peoples etc. Many users that express their opinion regarding any entity/Product, there may be chances that reviews are not written in correct form (Dictionary). Because of reviews available on these sites may contain noise such as spelling errors, typographical errors, standard abbreviations, and elegant writing. It is necessary to make data noise-free so that it can be used for opinion extraction. This paper describe a framework that was proposed to conduct opinion analysis of noisy reviews using techniques such as calculate similarity of terms and frequency of the document. The reviews of different products have been tested by this framework and the corresponding result is shown in negative (-ve) and positive (+ve) form. The results are satisfactory for all the tested products.

***Keywords*:** Opinion mining, Sentiment Analysis, Opinion extraction, Document Frequency

I. INTRODUCTION

In recent years, due to many research problems and practical applications, analysis of opinions (or analysis of feelings) has attracted great attention from researchers in the process of natural language and data mining [1]. It has also proved useful for companies, recommendation systems and publishing sites to create summaries of the experiences and opinions of people who consist of additional subjective impressions or even a positive or negative polarity of the review [2]. Data-based methods, which resist traditional text categorization techniques, are the classification of document polarity, which represents a significant challenge.

Opinion, we evaluate opinions and take decision. Same thing is applied to organizations when they introduce new product or on the way to introduce it; organizations take opinions of its customers in the form of reviews of product on official websites of organization, social media sites such as Facebook, Twitter, Blogs or online shopping sites. Customer also wants to know opinions of existing users before they use service or purchase a product. These reviews help organizations and its customers to evaluate the response or love among people about product or service [2]. The analysis of **Sentiment**, also called the opinion of miners, is the field of study that analyzes the opinions, feelings, evaluations, evaluations, attitudes and emotions of people towards entities such as products, services, organizations, individuals, problems, events, problems and their attributes [3]. It represents a large problematic space. There are also many names and slightly different tasks, such as the analysis of feeling, the extraction of opinion, the extraction of opinion, the extraction of opinion, the analysis of subjectivity, the analysis of emotions, the analysis of emotions, the analysis of opinions, etc [4].

This paper presents an unsupervised learning algorithm for classifying reviews. This algorithm takes text review or noisy reviews as input and gives output as whether the review is positive or negative and also handle the noisy error that is present in input review. For Handling the Erroneous in review in opinion mining there are many steps we prefer. This paper illustrates: **Section 2**, while presenting the related work that have done in this sector . **Section 3**, While presenting the Proposed Algorithm for handling erroneous review, it include following steps :- **1.** Firstly it extract the opinion phrases that containing **adjective or adverb**. As adverbs and adjectives are descriptors of another word and modify the meaning of word. These extracted words are called phrases. **Part of Speech (POS)** tagger is used for tagging the text (review). **2.** The second step is to assign **Semantic Orientation (SO)** of the extracted phrases. The phrase with positive semantic orientation has good association and the phrase with negative semantic orientation has bad association. **3.** Third step is to find whether the review is recommended (positive) or not recommended (negative). Finally calculate the average semantic orientation of phrases if it is positive then the review is classified as positive review (recommended) and if the average semantic orientation of the phrases is negative (non recommended) then the review is classified as negative review. **Section 4**, while presenting the Experimental Result and **Section 5**, while presenting the overall result , **Section 6,7** while presenting Future work and conclusion.

II. RELATED WORK

Product feature extraction, called also, opinion target identification, is crucial for opinion mining (OM) and summarization especially given that this task provides the foundation for opinion summarization [5]. The opinion target can be defined as the entity (i.e., person, object, feature, event or topic) about which the user expresses his opinion. Extensive approaches and techniques have been addressed to mine opinion components or targets from unstructured reviews. These works can be very broadly divided into two main categories supervised and unsupervised. Other works have also employed the semi-supervised approach. In the supervised learning approaches [7], a machine-learning model is trained on manually labeled data to extract and classify the feature set in the reviews. Although these techniques provide good results for opinion target extraction, they require extensive manual work for the training set preparation, they are also time consuming, and sometimes domain dependent. The most common techniques employed in supervised approaches are decision tree, support vector machine (SVM), K-nearest neighbor (KNN), Naïve Bayesian classifier and neural network. On the other hand, unsupervised approaches automatically extract product features using syntactic and contextual patterns without the need of labeled data .

The levels of document, sentence or even sentence (word) can usually be analyzed, opinions and feelings expressed in the revisions of the text. Extraction of opinions at document level (sentence level) is used to classify subjectivity or general sentiment expressed in an individual review document (sentence).

- Wei Jin and Hung Hay Ho [14] have proposed a new and solid approach to machine learning for web extraction and opinion. This model provides solutions for server problems that have not been provided by previous approaches. This system can self-learn new vocabularies based on the model it has learned, which is used in text and web mining. A new start-up approach is used to manage situations where the collection of a large training set can be expensive and difficult to achieve. In this paper, the effectiveness of the proposed approach in opinion extraction and the extraction of product opinion are demonstrated in the result.

- Guang Qiu, Bing Liu, Jiajun Bu and Chun Chen [15] focus on important opinion mining tasks which are the expansion of the vocabulary of opinion and the extraction of objectives. In this paper, a propagation approach to abstract the words and objectives of opinion in an iterative manner provides only a small initial vocabulary of opinion. The relationships identified between words of opinion and goals are used for extraction in this document. A new method is proposed to assign new words of opinion to the assignment of polarity and to the preparation of noisy publicity. The new approach is compared with others in the standard test data set. The result of this work shows that this approach works with other methods of last generation.

- Bo Pang and Lillian Lee [16] examine the relationship between the individuation of subjectivity and the classification of polarity. The detection of subjectivity may compress the revisions into shorter extracts that still retain the polarity information at a level comparable to that of the full revision. Using the Naive Bayes polarity classifier, it is shown that the subjectivity extract is a more effective voice than the source document. The document shows that the minimum cutoff frame leads to the development of an efficient algorithm for the analysis of feelings. Through this framework, contextual information can lead to a statistically significant improvement in the accuracy of the polarity classification.

- Niklas Jacob and Iryna Gurevych [17] show how a CRF-based approach to extract opinion goals in a single cross-domain environment. In this paper, a comparative evaluation of this approach is presented on the data set of four different domains. The performance-based CRF approaches a supervised reference line through the data set in the single domain configuration. The CRF-based approach also offers promising results in cross-domain configuration.

Opinion mining based on document, sentence, or phrase (word) level does not represent what exactly people like or dislike.

III. PROPOSED ALGORITHM

The main objective of this paper is to fetch reviews of a product of various companies and selecting the best product for the consumer by analyzing the reviews. A product is launched by various companies who provide different features for the same product in this work software is developed, such that it would find out the best product from various types by checking out the reviews available on the various social networking sites [18].

Opinion mining is the task to identify the user opinion about a particular object. So, the Opinion mining tool processes the reviews collected from different reviewers by generating a list of object/product features and performing aggregation of the opinions about each feature. Social networking sites and product selling sites are used by the user to express their opinion about products, events, people etc.

Unsupervised Review Classification:
In most cases, words and sentences of opinion are the dominant indicators for the classification of feelings. Therefore, the use of unsupervised learning based on such words and phrases would be very useful. This technique makes a classification based on some fixed syntactic sentences that will probably be used to express opinions [4,5] .

The algorithm consists of three steps:

Input: Written review
Output: Classification (i.e. positive or negative)
Step 1: Use part-of-speech tagger to identify opinion phrases.
Step 2: Estimate the semantic orientation of extracted phrase.
Step 3: Assign the given review to a class (either recommended or not recommended)

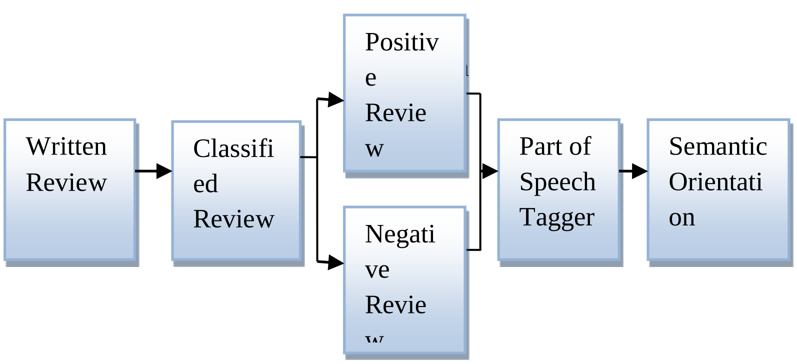


Fig. 1: Algorithm Steps Processing

All these steps are described in details –

3.1 Extract the Opinion phrases:

This is the first step of the process, in which from the posted review extract phrases that containing adjectives or adverbs. The reason for doing this is that research has shown that adjectives and adverbs are good indicators of subjectivity and opinions. So, the algorithm extracts two successive words, where one member of the pair is an adjective/adverb and the other is a context word [12].

- Two consecutive words are extracted from the review if they match with patterns in the table.
- Reason – **Adjectives & Adverbs** are good indicators of **Opinion**.

Single word adjective and adverbs may have different meaning in different context and they modify the meaning of other word quickly. For Example, the word "great" may have positive orientation in the movie review such as "great acting" and negative orientation in another movie review for "great loss". So rather than selecting single word adjective or adverb we selected bigrams containing adjective and adverb. Firstly POS tagging is applied to each word of review. Table shows POS tags and their meaning.

Table 1: POS Tags and Meaning

POS Tag	Meaning
JJ	Adjective
NN	Noun, singular
NNS	Noun, plural
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, present participle
VBN	Verb, past participle

Stanford POS Tagger

POS Tagger is used for tagging the text [19]

- The output is in the form of WORD/TAG eg. Nice/JJ – means "nice" is Adjective.
- Once the tagged text is available, find out the two consecutive words which have the tags mentioned in the above table.
- Below diagram describes the POS tagging process.

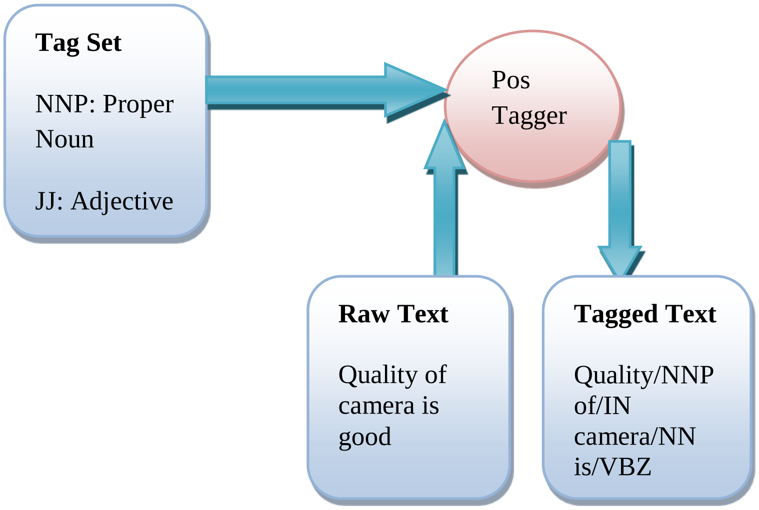


Fig. 2: POS Tagging

3.2 Estimate the semantic orientation of extracted phrase:

- Estimate the orientation of the extracted phrases using the **Point wise Mutual Information** (PMI) measure [20].
- PMI between 2 words, word1 and word2 can be defined as :

$$\text{PMI (word1, word2)} = \log 2 \left(\frac{\text{P(word1)\&word2)}}{\text{P (word1) P (word2)}} \right) \quad \ldots\ldots\ldots\ldots\ldots \text{(i)}$$

Here

- P(word-1 & word-2) = Probability that both words occurs together.
- P(word-1)*P(word-2) = Probability of co-occurrence of word1 and word 2, If both words are independent.
- $\frac{\text{P(word-1 \& word-2)}}{\text{P(word-1)*P(word-2)}}$ = Degree of statistical dependence between words.
- log = Gives information of presence of one word when we observe other.

The Semantic Orientation of a given review sentence or phrase is calculated by comparing its similarity to a positive reference word (“excellent”) with its similarity to a negative reference word (“poor”). More specifically, phrase is assigned a numerical rating by taking the mutual information between the given phrases and the word “excellent” and subtracting the mutual information between the given phrase and the word “poor”. In addition to determining the direction of the phrase’s semantic orientation (positive or negative, based on the sign of the rating), this numerical rating also indicates the strength of the semantic orientation (based on the magnitude of the number).

After calculation of PMI of phrase we calculate the Semantic Orientation of two-word phrase as given in. To find the **Semantic Orientation** (SO) measure of a phrase is calculated as follows:

SO(phrase) = **PMI** (phrase, “**excellent**”) – **PMI** (phrase, “**poor**”)

………………(ii)

When, the value of
SO is **+ve**: phrase is strongly associated with **excellent**.
SO is **–ve**: phrase is strongly associated with **poor.**

The probabilities are calculated by issuing queries to a search engine and collecting the are number of hits.

3.3 Assign the given review to a class:

The third step of analyzing the orientation of review is to take average of Semantic Orientation of each phrase of review and classify them as recommended or not recommended. If the average Semantic Orientation is Positive then review is classified as positive or recommended and if the average semantic orientation is negative then the review is classified as negative or not recommended.

3.4 Opinion mining process for Noisy reviews

In the previous work opinion mining of the error free reviews is described [4,21]. But, what happen if there are errors present in the reviews, in such cases we cannot identify nouns and adjectives easily, POS tagging would not work because the erroneous words are treated as Noun, for example –

Correct Sentence:
Quality of camera is good.

After POS tagging:
Quality/**NNP** of/IN camera/**NN** is/VBZ good/**JJ** ./.

Here the words {Quality, Camera} are tagged as Noun and {Good} is tagged as Adjective.
But if we consider the erroneous version of above text –

Erroneous Sentence:
Qlty of cam is gud.

After POS tagging:
Qlty/**NN** of/IN cam/**NN** is/VBZ gud/**NN** ./.

Observe that the words {Qlty, Cam} are tagged as Nouns because the tagger is not able to find these words in normal English words list, so they are treated as nouns. Also the word {gud} is treated as Noun, so we cannot identify the Features (Nouns) and Opinions (Adjective) directly in the given text. So, there is a need of a mechanism by which we can map these words correctly to features and opinions to make opinion mining possible on such erroneous reviews.

In this section we proposed a framework for mining the erroneous reviews which could handle the spelling mistakes and use of shortcuts in the review text. There are some steps that will be follow:-

Step 1: Stop word Removal
In this step the stop words are removed from the review, when using shortcuts the stop words are generally represented by single letters (e.g. “d” is written instead of “the”, “2” is written instead of “to”). Since we are not interested in the stop words, they are removed from the review.

Step 2: Dictionary Search
Each word Ei in the review is searched in the dictionary, to check if the word is in the correct form. If the word is found in the dictionary and it is an adjective then it is marked as opinion word.

Step 3: Feature List Lookup
When the word Ei is not found in the dictionary then or if the word is found in dictionary and it is noun then it is searched in the feature list and if exist then it is marked as feature. This step is required even if the word is found in dictionary because the features if a product might not be present in the dictionary.

Step 4: Opinion Dictionary and Feature Dictionary Lookup
When there is no direct match for Ei in dictionary or feature list, then we need to find a similar term present in the Feature list or Opinion list. It involves following steps.

Step 5: List Creation
As the number of opinions /feature terms is more, we create list of opinion/feature terms Ti having first character same as that of erroneous word and the edit distance between Ti and Ei is greater than a threshold. This step reduces the number of terms for which similarity should be calculated.

Step 6: Find Similar Terms
For each term Ti from above list, similarity is calculated using the formula described below – Similarity (W1, W2). If the similarity is above a predefined threshold then the term Ti is added in a set of matched terms.

Step 7: Add noisy term to Opinion/Feature list
Once similarity of all terms is calculated, the term (Tmax) with highest similarity value is considered as a match. If Tmax belongs to Feature list, then the erroneous word Ei is added to feature list as a noisy variation of the Feature. Similarly if Tmax belongs to Opinion list, then Ei is added as noisy variation of Opinion term Tmax.

E.g. - Suppose Ei= gud, and after calculating similarity we found that it matches with OPINION term Tmax =“good”, then we would add “gud” as a noisy variation of the Opinion “Good” in the opinion list. In future these noisy variations are used for lookup. E.g. In future if we encounter the word “gud” then by searching the list containing noisy variations of opinion, we can map “gud” directly to opinion “good” – without calculating similarity.

Step 8: Replace all occurrences of erroneous word
Once a match Tmax for Ei is found, all its occurrences in the current review and subsequent reviews are replaced by the Tmax, to make the processing faster.

Step 9: Similarity Calculation
Similarity between two words W1 (word from feature list / opinion list) and W2 (word from review) is calculated as –

$$\text{Similarity (W1, W2)} = \frac{\text{Length of LCS (W1,W2) x DF(W1)}}{\text{Length of W1 x LED(W1,W2)}} \qquad \text{………… (iii)}$$

where,
LCS = Longest Common Subsequence,
LED (W1, W2) = Levenshtein Distance between two terms,
DF (W1) = Document frequency of a term W1.

As the length of W1 (Word in Feature List / Opinion List) is always more than the erroneous version W2, therefore the length of LCS is divided by the length of W1. In the second part LED (W1, W2) indicates number of substitutions required to convert W1 into W2. We are interested in the Words which require fewer substitutions, so inverse of LED indicates that, lesser the edit distance, more similarity.

Finally the third part of expression DF (W1) – represents the frequency of occurrence of word W1 in the Error free Review dataset, we have assumed that the users generally use words (opinion and feature) which are more frequently used. E.g. Users generally speak about the features like {Flash, Picture Quality} and opinions {Good, Bad} more frequently.

Step 10: Determining Opinion Orientation
Once the opinion words and features are marked in the review, then the nearest opinion word is assigned to a feature in the review. And based on the average positive or negative orientation, we can conclude whether people like or dislike that particular feature.

Finding opinion orientation can be done using the Wordnet dictionary. Using Wordnet the orientation (positive or negative) is assigned to the opinion words). The process starts by using a seed set of opinion words and their orientation e.g. Good (opinion) and Positive (orientation), then the list of opinion is expanded by finding all synonyms of the opinion word, and all these of the opinion word, and all these synonyms are assigned the orientation same as that of the original word, similarly the antonyms of the opinion word are found using Wordnet dictionary and are assigned opposite orientation.

When an adjective is found in the review, first it is searched in the opinion list, if its present then opinion same orientation is assigned, else the all synonyms of the adjective are searched in the opinion list and if found then same orientation is assigned. Otherwise the antonyms of the adjectives are searched in the opinion list and if found then the opposite orientation is assigned to the adjective.

Step 11: Aggregation of Opinion
Once all the features, opinions and their orientation mentioned in the review are found out. Then for each feature number of positive and negative opinions is counted and if number of positive opinions is more than negative, then the opinion about the feature is considered as positive else it's marked as negative. All these steps are applied over a dataset which is known to have no error. The reviews, opinion words, features are indexed for further use. A opinion list is created which stores the opinion word and opinion orientation, also a features list indicating various features is maintained. The reviews are indexed for calculation of the document frequency as a part of similarity measure.

IV. Experimental Result Analysis

4.1 Noisy Reviews for Car:
In the given Table There is a sample of output when proposed algorithm is apply in given noisy review. The table shows two type output **Expected** means the output is driven by human and the **System detected** means the output is driven by system.

Table 2: Noisy Reviews for Car

Sn.	Review	Expected			System Detected		
		Feature	Opinion	Orientation	Feature	Opinion	Orientation
1.	pickup, I ffel 1.4 CRDi engn is hvng gud pickup fr Indian rods.	pickup engine	Good Good	+ +	Pickup Engn	Gud Gud	+ +
2.	Mileage of ths car is gud. I ffel 16/19 Kmpl In city/highway.	Mileage	Good	+	Mileage	Gud	+
3.	Safty fetures lik airbags,abs,collapsible stering Fetures lik automatic climate cntrl.	Safety	Like	+	Safty	Lik	-
4.	The acclrates systm is vry gud of i20 car. it caught 100 mrk in just few sec nd d gear system of d car is very smuth.	Gear	smooth	+	Gear	Smuth	_
		Accelerates	Good	+	Not found	Not found	---------
5.	D i20 car luks cerinly stylsh with its desgn, siting, boot nd engn vry nois free.	Engine	Noise free	+	Engn	Nois free	+
		Looks	Stylish	+	Not found	Not found	--------
6.	The sunruf was so stylish, Sinc dis is a new featurs, evrybdy wnts to have a luk onc it is opn. On the higways and in wintr seasn, it is vry plesurble.	Sunroof	Good	+	Sunruf	Gud	+
7.	Excllnt handlng and brakng systm suprb extrior nd clasy intriors,rumy passenger cabin.	handling	excellent	+	Handlng	Excllnt	+

		braking	excellent	+	Brakng	Excllnt	+
8.	I got 16.5 km milage rite frm d vry frst month isn't dat amzing. Gud cntrl, powerful headlites n brks mak ur jurney safe.	Mileage	Very	+	Milage	Vry	+
		brakes	powerful	+	Brks	Powerful	+
9.	My dzire gives me gud milage n ts a Powerful vcle in all terms.	Mileage	Good	+	Milage	Gud	+
10.	Luk nd Styl, New Model is relly impres2ive, d Desgn for Frnt Mud geard looks gud, and lamps Cmfrt.	Looks	Stylish	+	Luk	Styl	+

4.2 Result

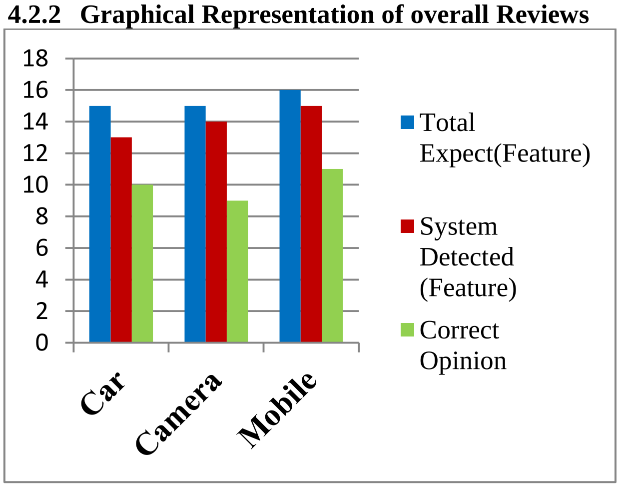
4.2.1 Statistical View

In Table, cumulative data of three product (Car, Camera, Mobile) has been aggregated in one data set, which will be used for calculating the accuracy and performance of our system based on our proposed methodology.

Table 3 : Overall Result

#Review	Total Expected (Feature)	System Detected (Feature)	Correct Opinion
Car	15	13	10
Camera	15	14	9
Mobile	16	15	11

4.2.2 Graphical Representation of overall Reviews



The above chart depicts the feature on the basis of opinions of each unit or class. From the above results we represent the performance of our proposed system in noisy review scenario. We analyzed the system accuracy on the basis of opinions of three items which are car, camera and mobile. Along with parameters i.e., the total expected and system detected feature. We have a parameter named correct opinion (which is exactly calculating the overall efficiency based on the accuracy of our system) i.e., the correct matches which our system made to achieve accuracy of feature extraction on expected level, respectively

4.3 Accuracy of Result

Accuracy defines the degree of closeness of quantitative measures to the actual value of this quantity.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{\text{TOTAL}}$$

where

TP = True positives is the measure of correctly classified positive reviews.

TN = True negatives is the measure of correctly classified negative reviews.

FP = False positives is the measure of wrongly classified positive review i.e. a negative review Wrongly classified as positive review.

FN = False negatives is the measure of wrongly classified negative review i.e. a positive review Wrongly classified as negative review.

Along with parameters i.e., the total expected and system detected feature. We have a parameter named correct opinion (which is exactly calculating the overall efficiency based on the accuracy of our system) i.e., the correct matches which our system made to achieve accuracy of feature extraction on expected level, respectively. With the help of given formula we are calculating accuracy of our system in two way:-

- Accuracy of System Detected Features = 87%
- Accuracy of System finding correct opinion = 74%

V. FUTURE WORK

In the current framework there is no way to distinguish between noisy/erroneous review and a correct review, so in future we will work on mechanism which would be able to differentiate them. For this purpose, we will append one more segregate function just above our proposed framework. This will help in reducing the time for redundant analysis of correct review. Also, the formula for calculating similarity can also be optimized further to improve the accuracy of the system. Thus as a future prospect, this classification will help in comparing two or more product based on their review submitted, the framework could be proposed for phonetic words (eg. photo - foto, kernel - colonel, seller – cellear, tea – t – tee, sea – c – see) and Neutral opinions, respectively.

VI. Conclusion

In this paper, we have represented a solution to take care of noisy reviews and formulated a similarity measure to identify a match between noisy word and the correct word. This framework can handle the opinion mining of reviews from social networking sites, where percentage of noise is more. The system has tested the ability to attain high accuracy and quality of sentiment prediction using the data harvested from a social network site. It includes the user's reception of opinions contained in the text and further improvements of the presented all expect to attain the improvement of classification performance due to the utilization of information derived from the social networks, namely, the information on relationships and connections between users. Firstly, we mentioned our problem statement and also some related work. Secondly, we discussed the related which have been done in the opinion mining and sentimental analysis sector. Finally, we proposed our framework for handling noisy data to get correct opinion. Our results predicting the correct opinion by our proposed system algorithm. Though segregation between correct review and noisy review could not be made out yet the overall performance depicted so far is very much profitable.

REFERANCES

[1]. B. Pang, and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, Volume. 2, 2008.

[2]. Anand Mahendran and Anjali Duraiswany, "Opinion Mining for text classification", International Journal of Scientific Engineering and Technology , Volume No.2, 2013.

[3]. Deepali Virmani, Vikrant malhotra and Ridhi tyagi, " Sentiment Analysis Using Collaborated Opinion Mining", 2014.

[4]. B. Liu, "Sentiment Analysis and Subjectivity," Handbook of Natural Language Processing, Second Edition, 2010.

[5]. M. Hu, and B. Liu, "Mining and Summarizing Customer Reviews", KDD' 04. USA, August 2004.

[6]. N. Jakob and I. Gurevych. Using anaphora resolution to improve opinion target identication in movie reviews. In Proceedings of the ACL 2010 Conference Short Papers, pages 263-268. Association for Computational Linguistics,2010.

[7]. J. S. Kessler and N. Nicolov. Targeting sentiment expressions through supervised ranking of linguistic configurations. In ICWSM. AAAI Press,2009.

[8]. Z. Toh and J. Su. Nlangp: supervised machine learning system for the classification of appearance categories and the extraction of opinion objectives. In the Proceedings of the 9th International SemEval Workshop.

[9]. M.Hu and B. Liu. Extraction and summary of customer reviews. In Proceedings of the 10th ACM SIGKDD International Conference on Discovering Knowledge and Data Mining, pages 168-177. ACM, 2004.

[10]. J. Jmal and R. Faiz. Customer review summary focus via twitter and sentwordnet. In Proceedings of the 3rd International Conference on Web Intelligence, Mining and Sementics , 2013.

[11]. L. Ferreira, N. Jakob and I. Gurevych. A comparative study of the characteristics extraction algorithms in customer reviews. In Semantic Computing, 2008 IEEE International Conference, in 2008.

[12]. A.M. Popescu and O. Etzion . "Extraction of product characteristics and review opinions. In the processing of natural language and text mining", 2007.

[13]. Forman, B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics.

[14]. W. Jin and H.H. Ho, "A Novel Lexicalized HMM-Based Learning Framework for Web Opinion Mining," Proc. 26th Ann. Int'l Conf. Machine Learning, pp. 465-472, 2009.

[15]. G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation," Computational Linguistics, vol. 37, pp. 9-27, 2011.

[16]. B. Liu. Mining opinion and analysis of sentiment. In Web Data Mining, pages 459-526. Springer, 2011.

[17]. N. Jakob and I. Gurevych, "Extracting Opinion Targets in a Single and Cross-Domain Setting with Conditional Random Fields," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 1035-1045, 2010.

[18]. M. Hu, and B. Liu, "Mining and Summarizing Customer Reviews", KDD' 04. USA, August 2004.

[19]. Phani gadde, L.V. subramaniam, Tanveer A. Faruquie, "Adapting a WSJ trained Part-of-Speech tagger to Noisy Text" Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data (J-MOCR-AND), 17-September 2011.

[20]. P. Turney, "Thumbs Up. Thumbs Down, Semantic Orientation Applied To Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), 2002.

[21]. L. Dey, and M. Haque, , "Opinion mining from noisy text data," Springer IJDAR , 2009.

[22]. Shikha parashar , Dr.R.K. Gupta, "A Survey on "Citizen Opinion Mining Initiatives In Government Decision Making", IJSART, ISSN 2395-1052, Volume -3, 7 –July 2017.

Authors Profile

Kirti Kushwah has completed M.Tech in the stream of Computer Science & Engineering from Madhav Institute of Technology & Science, Gwalior (M.P.). Her research area is based on Data Mining (Opinion Mining).

Rajendra Kumar Gupta has obtained Ph.D. from the ABV-IIITM Gwalior (M.P.), India. He is currently working as Professor in the Department of CSE & IT, Madhav Institute of Technology & Science, Gwalior(M.P.), India. His area of interest in new technology and concept applied in Data Mining Sector.

