IJCAT - International Journal of Computing and Technology
Volume 1, Issue 1, February 2014
www.IJCAT.org

Review of Online Product using Rule Based and Fuzzy Logic with Smiley's

[1]Amit Pimpalkar, [2]Tejashree Wandhe, [3] M. Swati Rao, [4] Minal Kene

[1, 2, 3, 4] Gurunanak Institute of Engineering and Technology, RTM Nagpur University,
Dahegaon, Kalmeshwar Road, Nagpur, Maharashtra, 441501, India

Abstract - Nowadays Internet has become a great means of communication in today's life there is a great development in the resources through which people can communicate such as discussion forums, online resources, groups and blogs. Internet encloses a bulk amount of subjective data which is scattered around. Opinion mining is used to identify and extract the subjective information from the internet with the help of data which are required for processing are natural language processing, text analysis. It is used to identify the attitude of the person. The person's opinion will help more while choosing any product while shopping. This will help to develop a system which will show the comments and reviews. We determine the polarity of sentiments of the person in the review and comments when the sentences occur in documentary level. It also identifies the different smiley which occurs in the comments and reviews. It compares the two different products with the help of reviews identified from the online resources which leads to find the best one of it. It uses Sentiwordnet and smiley's dictionary to determine the scores of each word present in the comment. Sentiments of words are classified in three scores, +ve, -ve and objective. It uses rule base and fuzzy logic approach to give the output.

Keywords **- SentiWordNet, Natural Language Processing, Sentiment Analysis System, Web Opinion Mining, Fuzzy measures, Text Tokenization.**

1. Introduction

Internet has become an essential need in every ones day to day life. It contains a large amount of textual data and is growing day by day. The abundant textual data is very easy to generate and publish on the web. People communicate through different online resources, groups, blogs, discussion forums, etc. it has become very hard task to extract the valuable information rather than to access it in a sequential manner. It takes very large time to gain the information manually and organizing it in a proper format. Organizing the text in a proper format and to distinguish is very hard task. Textual data on the web are of two types, facts and opinion. Facts focuses on the objective data transmission and opinion express the sentiment of the person. Facts can be expressed with the keywords but very hard to find the keyword from the opinions.

Sentiment analysis is the opinion mining used to identify the text on the web. It checks the attitude of the person, the judgments towards the particular statement, affective state or emotional communication. The main task of this system is to gather the more number of reviews from different online websites and overall detail of that particular product and then the polarity of the given text at document level sentences is checked. It confirms whether the contents of the documents are +ve, -ve or neutral. The highly existence of social media had made a lot interest in sentiment analysis. With the reviews, ratings of the online opinion will lead to identify new opportunities and manage their reputations. It helps the business persons to update their product if there any updating required. The sentiments may vary from person to person and are illogical.

The focus of the system is to analyze the sentiments for any product reviews. The input is to be taken from the product review web sites and the social networking sites on which the comments are posted for any particular product. Rule based method approach is used to properly filter the output for any product reviews. It also accepts the smiley's of the product so that it will help to directly identify the sentiment of the product. It also leads to comparison between two products, to identify the best product between the two.

2. Related Work

A. Pimpalkar [1] studies the role of sentence level statement, in which all type of sentences are used such as positive, negative, not only but also, etc. this type of statements were recognized. S. Kawathekar [2] studies the

role of negation in an opinion-oriented information-seeking system. They investigate the problem of determining the polarity of sentiments in movie reviews when negation words, like not and hardly arise in the sentences. A. Shukla [3] presented a tool which tells the quality of document or its usefulness based on annotations. Annotation can include comments, observation, notes, highlights, explanation, underline, question or help etc. Collective sentiments of annotators are classified as positive, negative, objectivity. A. Khan, B. Baharudin and K. Khan [4] presented a tool which gives the online review and which was a sentence level statement. C. Hauff [5] gives the way how to handle the negation words like not, no, neither, couldn't, etc words in the sentence. It may happen that even if the negative words are present in the sentence still its meaning is in a positive way. A. Joshi, Balamurali AR, P. Bhattacharyya and R. Mohanty [6] studied on the micro blogs, in which we can see and feel it reviews and gives the meaning of it. Animesh Kar, Deba Prasad Mandal [7] presented the tool in such a way that it gives the actual strength of the opinions and knows how the opinion is strong.

A. Neviarouskaya [8] performs fine grained categorization of sentences using ten categories: nine emotions ('Anger', 'Fear', 'Guilt', 'Disgust', 'Interest', 'Sadness', 'Joy', 'Shame' and 'Surprise') and neutral. The proposed rule-based approach processes each sentence in stages, including symbolic indication processing, detection and conversion of abbreviations; sentence parsing, smileys and word/phrase/sentence-level examination. Each analyzed sentence is automatically annotated with emotion or neutral label and numerical intensity; the strength of emotion. It also mentioned the links where the datasets would be available. B. Tierney [9] presents the results of applying the SentiWordNet lexical resource to the problem of automatic sentiment classification of film reviews. It comprises counting negative and positive term scores to determine sentiment direction is presented by building a data set of relevant features using SentiWordNet as basis, and correlated to a machine learning classifier. M. Thelwall [10] gives the hybrid knowledge for involving the Rule-Based and Support Vector Machines method. The hybrid approach gives the maximum efficiency for analysis of sentiments. Recent research has tried to automatically determine the "PN-polarity" of subjective terms. F. Sebastiani [11] in order to aid the extraction of opinions from text i.e. identifies whether a term that is a marker of opinionated content has a positive or a negative connotation. P. Bhattacharyya [12] proposed a technique for the effective SA of movie reviews. It also describe a novel approach to process the predictions for individual documents of the test dataset to improve the accuracy over the entire set. It presented a WorldNet based method for the effective incorporation of linguistic information in the system without any kind of experts' intervention. It also presents a generic method that can be used to improve the accuracy of classification over a test dataset in any kind of classification task. It shows how the application of this technique to SA helps to attain the best accuracy so far in this field. B. Pang [13] gives the idea about the Support Vector Machines technique related to sentiments analysis. It also includes the other technique like Maximum Entropy.

3. Research Gap Found

- Parsing: the identification of subject and object in the sentences which actually refers to verb or adjective is very hard.
- Identification of entities – Involves the resolution of referential expressions such that each entity mentioned in a summary can easily identified by the reader.
- More complex sentences**:** Any piece of text consisting of more than one sentence. Concerns the relationships between the sentences in a summary, as well as those between individual sentences and the overall summary.
- Anaphora resolution: we went for drive and had dinner; it was terrible. It is the problem of resolving what a noun or pronoun refers to.
- Discourse**:** Any piece of text consisting of more than one sentence. Concerns the relationships between the sentences in a summary, as well as those between individual sentences and the overall summary

4. Proposed System

It collects all the customer reviews which contain the facts and opinions. The sentences are split into subjective and objective ones based on the dictionary. Subjective sentences are further processed for extraction to categorize as +ve, -ve or neutral opinions. A rule based method is used for the categorization of subjective and objective sentences. From the subjective sentences, opinion expressions are mined and their semantic scores are checked using the SentiWordNet and smileys directory. The final score of each individual sentence is calculated after considering the whole sentence structure, contextual information. The steps are,

- Splitting the reviews into sentences and make a Bag of Sentences.
- Then Remove noise form sentences using spelling correction, stop words convert special characters and

symbols to their text expression, use Part of Speech for tagging each word of the sentence and store the position of each word in the sentence.

- Make a dictionary of the important feature with its position in the sentence.
- Calculation of assigned score is done to obtain the final result of the product.

4.1 Sentiwordnet

SentiWordNet is a lexical resource for the opinion mining. SentiWordNet assigns to each word of three sentiment scores: positivity, negativity, objectivity. A triplet of numerical scores Φ (p, s) for p ∈ {P, N, O} describing how strongly the terms contained in Sentence (s) enjoy each of the three properties. ∑ {P, N, O} =1

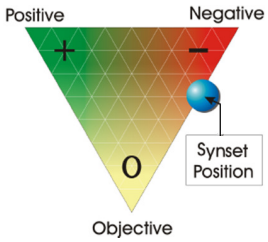


Figure 1: Graphical Representation of SentiWordNet

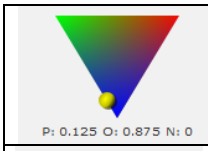
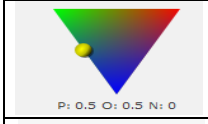
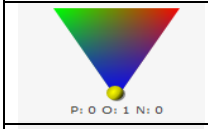
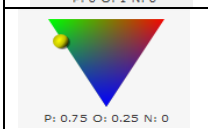
	Used as verb: Best#1; “the book was best to buy”
	Used as adverb: Best#1; "she played best after a couple of weeks"
	Used as adverb: Best#2; "you'd best stay at college"
	Used as adverb: Best#3; "he knows best"
	Used as adjective: Best#1; "the best reply"; "the best time for studying"
	Used as adjective: Best#2; "the thought was best not to respond”

Figure 2: Representation of words in SentiWordNet

The above Figure 1 shows the graphical representation of the SentiwordNet. It represents the opinion of any particular word. Opinions of the word will lead us to know the score of that word. The summation of the any word will give us the result 1. It is the addition of +ve , -ve and objective score.

Figure 2 shows the example of SentiwordNet to see how how the values are assing to particular word. For example, best is the word and is classified into adjective, verb, adverb, noun,etc. All different types of best word will give different values.

4.2 Phases of the System

1) Data collection: Data collection is the process of gathering and measuring the information from different resources of the interest in a systematic fashion that will enables us to answer the questions, test hypotheses, and evaluate the result. The data collection will be done from the reviews of the customers.

2) Identification of data: It identifies the data according to its value and what we are going to use. After data has been identified it will be given as an input to the system.

3) Pre-processing: The pre-processing is done to remove the unnecessary words or irrelevant words from the customer’s opinion. Our system deals only with the description of part of speech each review, processing means breaking review into words to differentiate it.

4) Part of Speech tagging: it parses each sentence and yields the part- of speech tag of each word whether the word is a noun, verb, adjective, adverb, etc and identifies simple noun and verb groups. If one sentence contains no frequent feature but one or more opinion words, find the nearest noun or noun phrase of the opinion word as an infrequent feature. Speech tagging is used to identify every word of feedback as either noun, adjective or adverb

5) Negation detection: It is also an important element of implementing sentiment analysis by using term scores, since negation in a sentence such as “I did not find this product funny or interesting” would invert the opinion orientation of otherwise positive terms such as “funny” and “interesting”.

6) Stop Words Removal: We eliminate words like prepositions, digits, articles and nouns like name of product etc with the help of Parts of speech tagging method as their existence are meaningless in system. It assists better extraction of opinion phrases/words from the tagged file.

7) Rule-Based approach: In the Rule-Based approach, rules are to be defined which contains some defined relation which have originator and its associated

resultant. In this methodology, certain rules are to be defined and then the sentiments should be viewed or analyzed depending on rules.

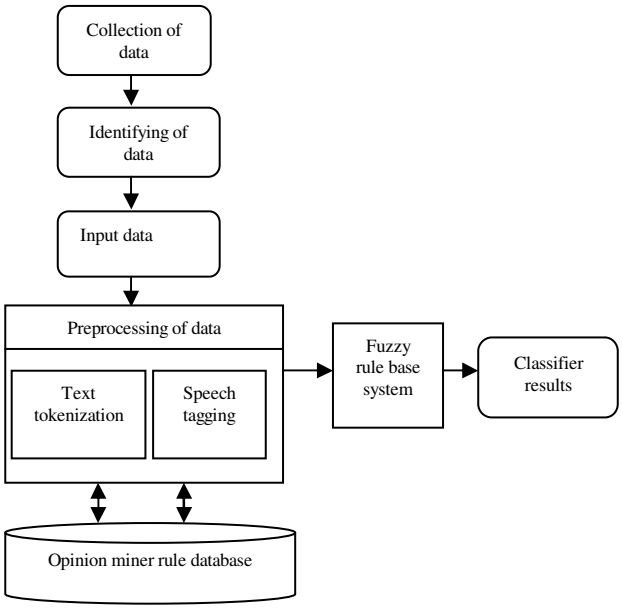


Figure 3: Sentiment Analysis System

This system works when the comparison of the two products has to be done. It takes the link of the particular product from where has to be taken and the category of that product. It then collects all the reviews or comments and image from the online resources such as flipkart.com, snapdeal.com, ebay.com, etc. It shows the overall specification of that product and the maximum reviews are displayed. The comparison of any two products can be done with the help of reviews. At last the sentiment score is calculated and the best product is displayed.

4.3 Opinion Analyzer

Opinion analyzer obtains the collective sentiments of annotators i.e. the statements. There are two algorithms from that we judge the quality or usefulness of documents. In first algorithm, it extracts all the annotations created by annotators which are of type underline, highlights, and comments and count total number of annotations present on the document. In second algorithm, it computes collective sentiment of annotators.

Algorithm 1 Find the total count of annotation found in the document.
Input: Annotated document
Output: Total Count of Annotation present in the Document
1. **for** each annotated document **do**

$$\text{Count} = \sum_{I=0}^{N} \text{Annotation} \quad (1)$$

2. **End for.**
3. Output Count.

Algorithm 2 Find the average score of each annotation found in document
Input: List of sentiment words extracted from comments of annotation
Output: Sentiment score and Sentiment Review

1. **for** each sentiment word from List **do**
   Get polarity as well as sentiment scores using *SentiWordNet*.
2. Compute the intensity of each word using Fuzzy Intensity Finder algorithm.
3. Final Sentiment Score (SS)

$$SS = \sum_{I=0}^{N} \frac{\text{Max(Polarity)}}{\text{Count}} \quad (2)$$

4. Output Sentiment Score.
5. Output Overall Sentiment Review.

4.4 Algorithm Fuzzy Intensity Finder

We calculate the weight of the extracted opinionated phrases as the weights of individual words in the phrases.

Case 1: There are few adverbs like very, really, extremely, simply, always, never, not, absolutely, highly, overall, truly, too, etc. (we imply 15 such adverbs) which may be used positively or negatively like {very good, very bad}, {extremely acceptable, extremely unacceptable}, {too good, too bad}, {simply outstanding, simply disgusting} etc.
Case 2: Again never, not etc. will change the orientation of the opinion like {not good will be bad} {never accepted will be always rejected} etc.
Case 3: Case 1 and Case 2 may also come together in opinion phrases like {not very good}, {not absolutely recommended}, {not truly reliable one} etc. We consider them accordingly in the calculation of fuzzy weights of opinion phrases.
Case 4: It is also possible that in some cases like {It is not only good but also it's an awesome} etc. appears in sentiments.

In Case 1, we consider the weight (W) of the opinion like (very /extremely/highly etc) (Adj)

$$= \sqrt{\text{Value Of (Adj)}} \quad \text{if ValueOf(Adj)} \geq 0.5 \quad (3)$$

$= (Value\ Of\ (Adj))^2$ if ValueOf(Adj) < 0.5 (4)

Table 1: Fuzzy Measure of (Adverb, Adjective) Phrases

Great	1.0	Simply Great	1.0
Fabulous	0.75	Simply Fabulous	0.5625
Horrible	0.375	Very Horrible	0.1406
Strange	0.625	Very Strange	0.3906

In Case 2, we consider the weight (W) of the opinion like (not/never)(Adj /Verb)
= 1 - ValueOf(Adj/Verb) (6)

Table 2: Fuzzy Measure of (Not/Never, Adjective) Phrases

Horrible	0.3750	Not Horrible	0.625
Pathetic	0.375	Not Pathetic	0.625
Conclusive	1.0	Never Conclusive	0.0
Stunning	0.75	Never Stunning	0.25

In case 3, we consider the weight (W) of the opinion like (not/never) (very /extremely/highly etc)(Adj)
$= \sqrt{(A * B)}$ (7)
Where A = (very/extremely/highly etc) (Adj)
$= \sqrt{Value\ Of\ (Adj)}$ if ValueOf(Adj) ≥ 0.5 (8)
$= (Value\ Of\ (Adj))^2$ if ValueOf(Adj) < 0.5 (9)
And B = (not/never) (Adj)

Table 3: Fuzzy Measure of (Not, Adverb, Adjective) Phrases

Fabulous	0.750	Not Fabulous	0.250	Not Very Fabulous	0.375
Horrible	0.375	Not Horrible	0.625	Not Very Horrible	0.2965

In case 4, we consider the weight (W) of the opinion like (but/also/nor) words. Firstly we divide the sentiment into two opinions group. Then these two opinions separately calculated as follows. Finally the average of both cases will be taken as a Fuzzy measure for the given sentence.

F= $\sqrt{(\sqrt{A} + \sqrt{B})/2}$ if ValueOf(Any Adj) ≥ 0.5 (10)

$F = ((A^2 + B^2)^2)/2$ if ValueOf(Both Adj) < 0.5 (11)

Table 4: Fuzzy Measure of (But, Also) Phrases

Good	0.625	Fantastic	0.3750
$\sqrt{Good}$	0.7906	$\sqrt{Fantastic}$	0.6123
This is a not only good but also fantastic			0.8375
Bad	0.25	Horrible	0.375
Bad^2	0.0625	$Horrible^2$	0.1406
This is not only bad but it's horrible			0.0206

5. Conclusion

Web contains the social networking sites such as, twitter, Facebook, orkut, etc. and other such sources obtain the opinion from the user about any subject and help to answer the question what people are interested in. It will help the companies to properly re-design the product if any problem or drawback has occurred in it. It will also check all type of smiley's because most of the people uses smiley to express their emotions.

References

[1] Amit Pimpalkar, "A System for Sentimental Analysis of Movie Reviews Involving Rule-Based and Fuzzy Measure", International Journal of Artificial Intelligence and Knowledge Discovery (IJAIKD), ISSN 2231-0312, Vol.3, No.2, 2013.

[2] Swati A. Kawathekar and Dr. Manali M. Kshirsagar, "Movie Review analysis using Rule-Based & Support Vector Machines methods", IOSR Journal of Engineering, Vol. 2(3) March 2012, pp. 389-391.

[3] Archana Shukla, "Sentiment Analysis of Document Based On Annotation", International Journal of Web & Semantic Technology (IJWesT) Vol.2, No.4, October 2011 pp. 91-103.

[4] Aurangzeb Khan, Baharum Baharudin and Khairullah Khan, "Sentiment Classification Using Sentence-level Lexical Based Semantic Orientation of Online Reviews", Trends in Applied Sciences Research, Vol. 6, July 2011, pp. 1141-1157.

[5] C. Hauff, Dadvar, Maral and Jong de, Franciska, "Scope of negation detection in sentiment analysis", Dutch-Belgian Information Retrieval Workshop, Netherlands, February 2011.

[6] Aditya Joshi, Balamurali AR, Pushpak Bhattacharyya and Rajat Mohanty, "C-Feel-It: A Sentiment Analyzer for Micro-blogs", The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, June, 2011, pp. 12.

[7] Animesh Kar, Deba Prasad Mandal, "Finding Opinion Strength Using Fuzzy Logic on Web Reviews", International Journal of Engineering and Industries, volume 2, Number 1, pp 37-43, March, 2011.

[8] Alena Neviarouskaya, Helmut Prendinger, Mitsuru Ishizuka, " Model: novel rule-based approach to affect sensing from text", Natural Language Engineering, Cambridge University, Vol. 17, pp. 95- 135, September 2010.

[9] Brendan Tierney and Bruno Ohana, "Sentiment Classification of Reviews using SentiWordNet", 9th IT&T Conference, Dublin Institute of Technology, Ireland, October, 2009.

[10] Mike Thelwall and Rudy Prabowo, "Sentiment Analysis: A Combined Approach", Journal of Informatics, School of Computing and Information Technology, University of

Wolverhampton, UK, Volume 3, Issue 2, pp. 143-157, April 2009.

[11] Fabrizio Sebastiani and Andrea Esuli, SENTIWORDNET: "A Publicly Available Lexical Resource for Opinion Mining", Proceedings of the 5th Conference on Language Resources and Evaluation, Genoa – Italy, pp. 417-422, May 2006.

[12] Pushpak Bhattacharyya, Akhel Agrawal, "Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be Classified", International Conference on Natural Language Processing, IIT Kanpur, India, December, 2005.

[13] Bo Pang, Lillian Lee, and S. K. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing, Volume 10, pp. 79-86, 2002.

First Author has obtained his bachelor degree in Computer Engineering from Nagpur University in 2005 with first class and master degree with Honors in Computer Science & Engineering with specialization in Software Systems form RGPV, Jabalpur in 2013. He is currently working as Assistant Professor in Gurunanak Institute of Engineering & Technology, Nagpur. He published one international Journal paper and presented paper in two international conferences and five National conferences. His current research interests Data mining, Computer Graphics, Data structure, and Database.