An Implementation of Opinion Mining Using Fuzzy Inference System

Rashmi Priyadarshi
M.Tech(CS&E)
Amity University Noida, India
singhrashmipriya@gmail.com

Seema Rawat
Assistant Professor
Amity University Noida, India
srawat1@amity.edu

Praveen kumar
Assistant Professor
Amity University Noida, India
pkumar3@amity.edu

Abstract— **Opinion mining is a technique by which we can find the semantic strength of reviews on an object using Fuzzy Inference System. Today people use information space to broadcast their opinions or views on any object or situation across the world. Analyses of these views yield a very useful information i.e. overall thinking about any object or situation. In this paper we have taken data from social networking site facebook on samsung mobile phones for analysis. After applying FIS on data collected we will give overall view on Samsung mobile phones across the virtual reality.**

Keywords— **Opinion mining, FIS**

I. INTRODUCTION

Now a day researches are going on in field of mining of views on any object or any situation that is known as semantic analysis. It is very interesting and useful to extract useful information from English or any other natural language automatically. Automatic knowledge discovery is very convenient because it will not take so much to read the whole views. It will also help in not getting confuse. There are so many methods for opinion mining which are as follows:

(i)Unsupervised – e.g. NLP pattern
(ii) Supervised – e.g. SVM, Naïve Bays
(iii) Semi-Supervised- e.g. lexicon + classifier

And in this report Fuzzy Inference system is used for opinion mining. FIS is also the most dynamic research area in the application of fuzzy set theory, fuzzy reasoning and fuzzy logic. It has a wide range of applications in numerous industrial and commercial merchandises and systems. Several applications are associated to nonlinear, time-dependent, vague systems and also complex systems. FL (fuzzy logic) begins with the idea of fuzzy set. Set which don't have any crisp boundary i.e. a set of elements which has partial values also. As in our world maximum things not have sharp boundary.

For example: A Jug half filled with water can also be said to half empty. So it has membership value in both set i.e. empty and filled. It is not bounded to anyone but lie in both. There are many examples like this, take Friday. Friday is weekday or weekend? We feel like both kind of day for Friday, therefore it has some value of membership in weekday and some value of membership in weekend.

A. OPINION MINING:

Opinion mining is a technique by which analysis of attitude of speakers is done about any object on which they have given their views. Using this analysis a consolidated view about the object is created. For opinion mining subjective information is used i.e. personal opinion, interpretations, points of view, emotions and judgement. We know that researches are going on in this field as opinion mining is very helpful for the company, project developers and customers. Information for opinion mining is gathered in two ways:

(i) Pre-web:

In this information is gathered from friends, relatives, acquaintances and consumers report.

(ii) Post-web:

In this information is gathered from blogs, e-commerce sites, review sites and discussion forums. When information got gathered semantic analysis of that is done. But this is not so easy as it is for human brain. Some comments are well structured but some comments which are posted by general consumers are unstructured. For example:

"The Samsung phones are not good"(structured sentence).

"wow@samsung"(unstructured).

Not only different kinds of sentences make it difficult for mining but today people use spoken words for writing, smiles, special characters etc. are also use in abundance and enhancing the difficulty level for mining. Besides the difficulties for analysis of these kinds of words/sentences different techniques used for opinion mining. So to get useful knowledge from the structured and unstructured or any kind of comments opinion mining is done.

(1) Steps use for opinion mining[11] here are as follows:

(i) Gathered the relevant comments related to product.

This is done by surfing the internet finding reviews on twitter, facebook related to product then crawl it down and picks relevant sentences.

(ii) Find the opinion word and its semantic orientation.

In this step, words will be saved in database and then which words are negative and which ones are positive is to be found.

©2014 CIPECH14.

(iii) Check link between product characteristics and its semantic orientation.

In this step, what is the feature and related to that what is the semantic orientation of the word is to be found.

(iv)Find strength of word and depending on this overall view on product is made.

In this step, the value of negativity and positivity of the words is calculated then these words are analyse and over all perception about the product is made.

(B) Basic components of Opinion Mining:

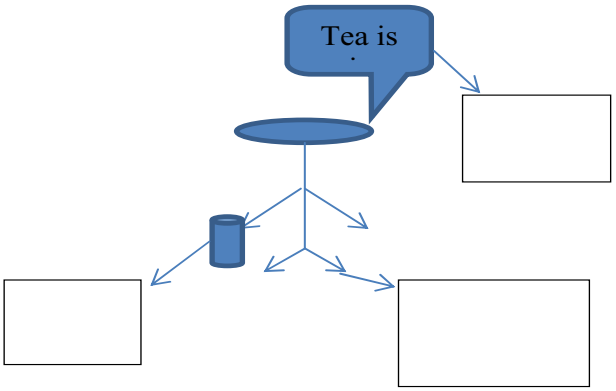

Fig1: Basic Components of Opinion Mining.

B. FUZZY LOGIC:

(1). Fuzzy logic introduced in the year 1965 by Lotif A. Zadeh. It is a mathematical tool for dealing with uncertainty. Dr. Zadeh states that the principle of complexity and imprecision are correlated i.e. "the closer one looks at real world problem, the fuzzier becomes its solution [4]." The fuzzy logic provides a mechanism for representing linguistic constructs such as "high", "low", "warm", "tall" etc. In general, fuzzy logic provides an inference structure that enables appropriate human reasoning capabilities [5].
In recent years fuzzy logic implementation area has grown at very fast rate i.e. is fuzzy logic methodology used to built washing machines, microwaves, cameras, medical instrumentation, camcorders and decision-support systems etc.
In fuzzy logic work is done on fuzzy sets. Fuzzy set is collection of membership value between 0 and 1[4].
Some of the operation that can be done on fuzzy sets are as follows:
(i) Union:
$\mu_{A\cup B} = \max [\mu_{A(x)}, \mu_{B(x)}]$ for all x€ U

(ii) Intersection:
$\mu_{A\cap B} = \min [\mu_{A(x)}, \mu_{B(x)}]$ for all x€ U
(iii) Complement:
$\mu_{\bar{A}(x)} = 1 - \mu_{A(x)}$ for all x€ U

(iv) Algebraic Sum:
$\mu_{A+B} = \mu_{A(x)} + \mu_{B(x)} - \mu_{A(x)}.\mu_{B(x)}$

(v) Algebraic Product:
$\mu_{A.B(x)} = \mu_{A(x)} \cdot \mu_{B(x)}$

(2) Some of the properties[4] that fuzzy sets holds are as follows:

(i) Commutativity
(ii) Associativity
(iii) Distributivity
(iv) Idempotent
(v) Identity
(vi) Involution
(vii) Transitivity
(viii) De Morgan's Law

(3). Fuzzy Expert System/Fuzzy Inference System
It is also known as rule-based systems, fuzzy models. This system is for decision making. It use If-Then rules for mapping input into output using connectors "OR" or "AND". The connector "OR" stands for 'max' while "AND" stands for 'min'. We can custom these also. Below figure shows FIS structure:

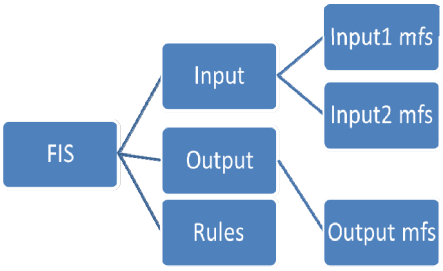

Fig2: FIS Structure.

(a) The FIS is constructed of five functional blocks:

(i) A rule base that contain rules.
(ii) A database that defines the membership functions of fuzzy set.
(iii) Decision-making unit that perform operations on the rules
(iv) Fuzzification Unit that converts the crisp quantities into fuzzy quantities.
(v) Defuzzification Unit that converts the fuzzy quantities into crisp quantities.
There are two types of FIS that uses for solving problem i.e. Mamdani and Sugeno FIS. The main difference between Mamdani and sugeno method is that output membership functions generated by sugeno are either linear or constant [6]. Enumerating membership value of elements in fuzzy set is done by defining membership function or graphically as shown below for example:

$\mu_{highoil\text{-}price} = \{5/0+\ldots.+16/0.15+\ldots.+20/0.85\}$

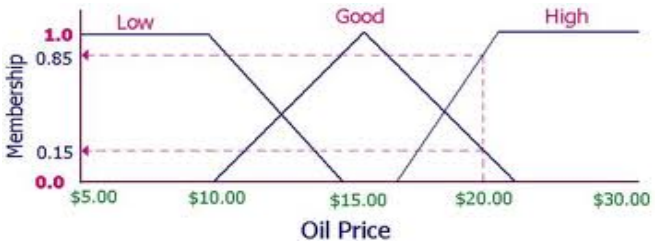

Fig3:Membership Function.

II. MATERIAL AND METHODS:

(A). Data Collections:

Reviews are collected from facebook on Samsung mobile phones. Due to time as constrain a few amount of data is collected nearly hundred sentences.

(B). Constructing opinion word with their score:

A pool of words relevant to Samsung mobile phones is created. Some of words are negative and some of words are positive but with the help of sentiwordnet, scores for words is found i.e. positive, negative and neutral also [10]. I collected words with the help of sentiwordnet which is having positive and negative score greater than zero [10]. Assigned '-'sign to negative words and '+' sing to positive words. Scores of some words are represented on number line drawn below.

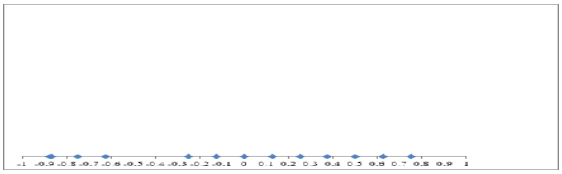

Fig4: Number Line showing negative, neutral and positive words with their scores.

(C) Experimental structure:

For data collection, php code has been use for crawling and for preprocessing java script has been used and for drawing result and analysis fuzzy Inference system of Matlab 2009a has been used. Then drawn value has been aggregated by taking sum.

III. IMPLEMENTATION OF THE APPROACH

As we know there is vagueness in natural language and fuzzy logic work on vagueness. In fuzzy logic we can take linguistic variables and draw their numeric value [5][7]. Here same is done using FIS triangular function. We define two linguistic variables for words which has been collected i.e negative and positive. The input value lying from 0 to 1 having these linguistic variables map to output value lying from 0-1 having linguistic variables negative, neutral and positive. Creation of input and output functions and variables of FIS is shown below with matlab snapshot.

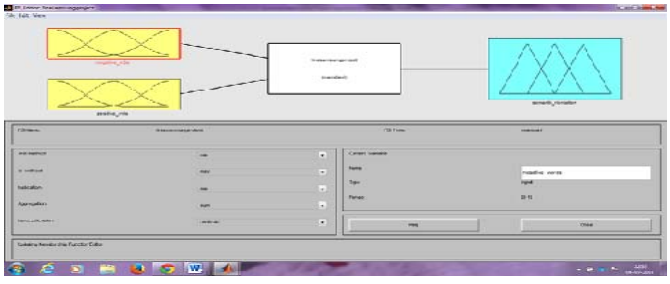

Fig5: FIS Editor

A. Fuzzy Set*:*

The fuzzy set has been created by using opinion words. There are two fuzzy sets for the words i.e. $\mu_{(-)}$ represents the negative fuzzy set and $\mu_{(+)}$ represents the positive fuzzy set. Negative fuzzy set contain negative membership value of the opinion words and positive fuzzy set contain positive membership value of the opinion words.

Membership values of fuzzy sets found with the help of triangular function of FIS. According to value on x-axis of function, y-axis i.e. membership value for that value of x is found. Fuzzy set i.e. negative fuzzy set and positive fuzzy set with their membership values are shown below respectively and snapshot for that of matlab is also shown:

$\mu_{(-)}$={Best/0+…+cute/0+…+like/0+…+crash/0.125+…+ bad/0.67+…+warm/0.78+……+faulty/0.78+…+muck/0.89 +…+ poor/0.91}.

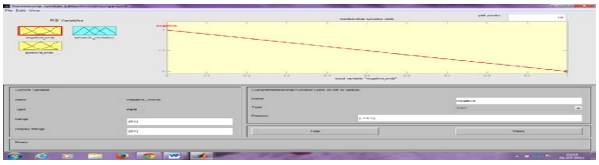

Fig6:Membership Function Editor showing Negative_word membership function.

$\mu_{(+)}$={poor/0+…+pathetic/0+…+fool/0+….+best/0.75+…+ lovely/0.63+…+wow/0.38+…+ easy/0.13}

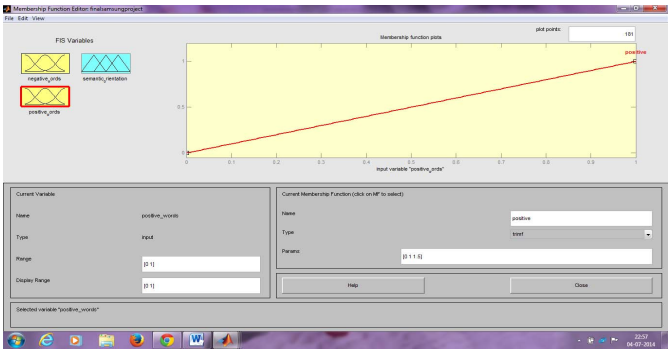

Fig7:Membership Function Editor showing Positive_words membership function.

B. Results and Discussion:

To get desired result rules for FIS is to be formed according to that result comes and analysis can be done. As we know that general way of representing human knowledge is by forming natural language expressions given as follows:

"IF antecedent THEN consequent", It is known as IF-THEN rule-based form.

The rule which we have form for getting desire result in Matlab is shown below as Matlab snapshot.

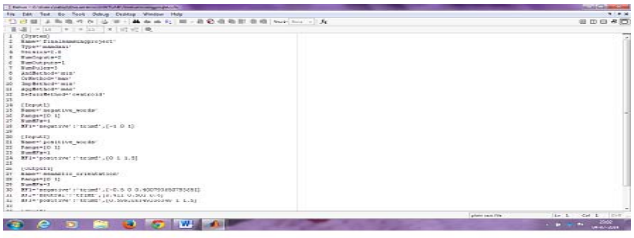

Fig 4.1 Editor showing Rules.

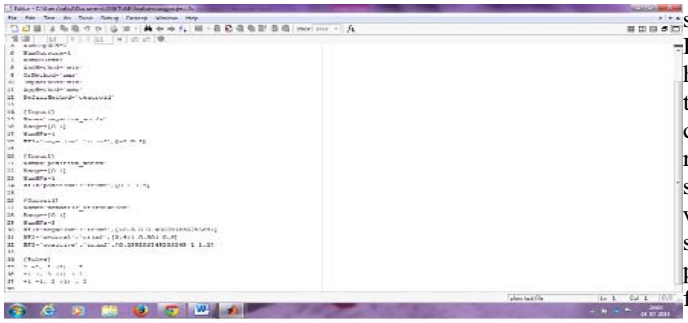

Fig 4.2 Editor showing Rules

At the end of above snapshot rules are represented in symbolic form. It is language neutral but no difference in display of rules. Machines deals with this kind of rules. The first column shows input variables, second column shows output variables, third column shows weight applied to rules and fourth column shows which connector is used if OR(2) or AND(1). The membership function's index number is shown by number of first two columns.

Below snapshot of matlap showing output variable. This also show how input variables are map to output variables.

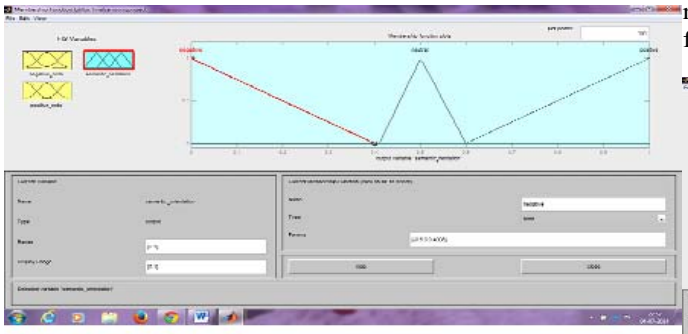

Fig8: Membership Function Editor.

Left corner is place for entering range in which we want to map our input to output function and at right hand corner we entered function name and its range on x-axis.

C. RULE VIEWER:

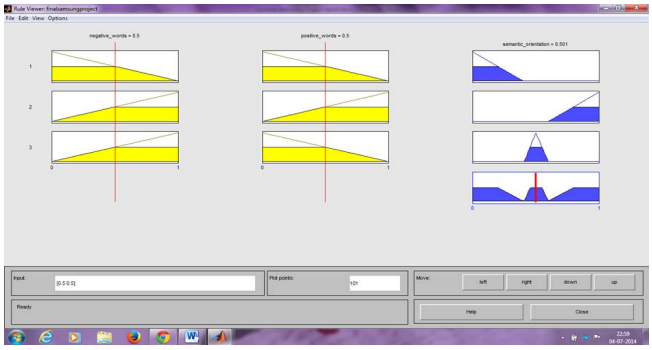

Fig9: Ruler Viewer.

Above snapshot is showing rule view and known as Rule Viewer. This shows roadmap of complete fuzzy inference system. It is based on FI snapshots explained previously. Here 10 plots are nested in this snapshot. The three horizontal plots at top show antecedent and consequent of the first rules. Row represents each rule of the system and column represents variables that are used. On left of each row rule number is shown. The first two columns of snapshot show membership functions by the antecedent or we can say 'if' part of each and every rule and last column show membership function referenced by the consequent part of the rule i.e. 'then' part. In the third column the fourth plot show aggregated weighted decision for the given inference system. The bold vertical line in fourth plot show defuzzified output.

The variables and their current values are displayed on top of the columns. In the lower left, there is a text field Input in which you can enter specific input values. To see working of whole fuzzy system we can change input value in left corner of rule viewer and get different result regarding our input value set. This is shown by following snapshots:

When negative_words value=0.125, positive_words=0 then semantic orientation=0.326.This show symentic orientation of view having these value for negative and positive is negative as we have use 'or' method so maximum value from negative and positive is taken

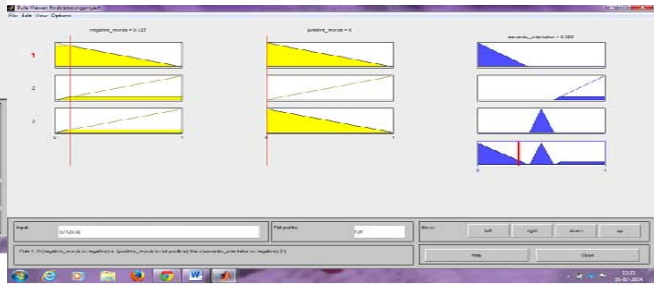

Fig10:Showing semantic_orientation of negative word.

Interpretation for the below snapshot is when negative_words value=0.875, positive_words=0 then semantic orientation=0.497.

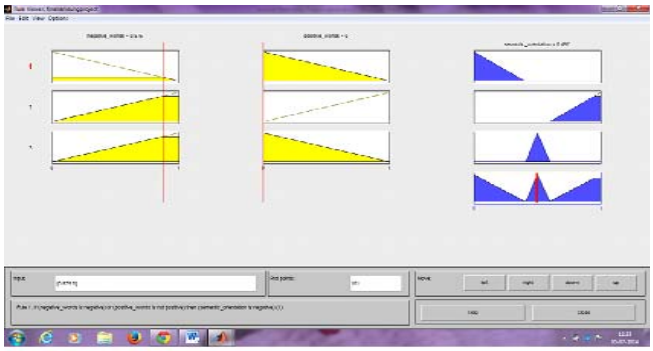

Fig11:Showing Semantic_orientation of Negative Word.

Below snashot is given for having negative_words value = 0.375, positive_words=0.375 then semantic orientation=0.447.

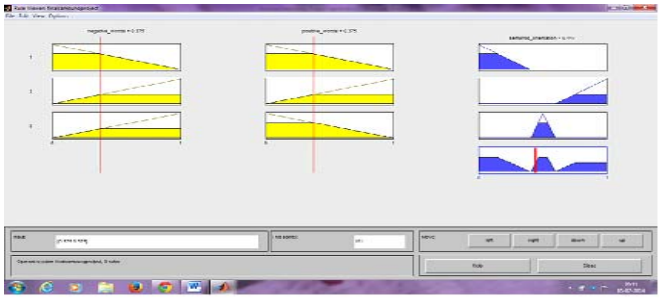

Fig12:Showing Semantic Orientation Of word having negative and positie value both.

Here we see a word having same value in negative and in positive but practicaly this can't happen. So this kind of word is neither taken as negative nor positive so work as neutral word. As impact is neutral.

Below snashot is given for having negative_words value = 0.0, positive_words=0.75 then semantic orientation=0.486.

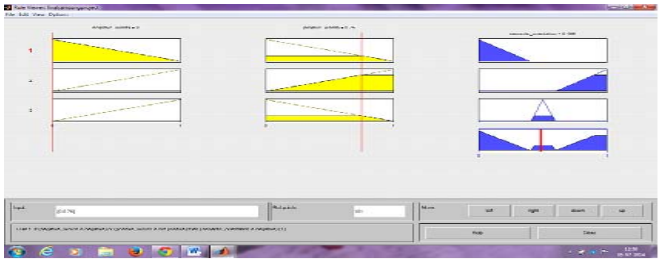

Fig13: Showing Semantic Orientation of Positive Word.

VI. CONCLUSIONS AND FUTURE ASPECTS

In this work views from social networking site facebook is collected. These views are both structured and unstructured. So for finding the strength of opinions expressed by reviewers on object, data is pre-processed and the relevant words are taken from the comments. Words are characterise as positive and negative by giving negative sign in front of value of negative word. Then these words are shown on number line. The negative words are shown on scale less than zero and positive on scale greater than zero. Then normalisation of the negative values on scale between 0-1 has been done so that inputs can be easily mapped to output membership functions in FIS. Two input fuzzy sets has been taken in FIS i.e. negative fuzzy set which shows negative membership value of the relevant words and positive fuzzy set which shows positive membership value of the relevant words 'μ(-)' symbol is given to negative fuzzy for representation and μ(+) symbol is given to positive fuzzy set for representation. After creating these input fuzzy sets we have mapped these inputs sets to single output fuzzy sets having three linguistic variables negative, positive and neutral. Negative shows the negative_orientation strength, positive shows the positive_orientation strength and neutral use to show the neutral_orientation strength. Input to Output mapping is done by creation of IF-THEN rule using OR as connector. Weight 1 is given. The formed rules are helpful in decision making. Decision making is a very essential societal, cost-effective and scientific effort. These activities involve the stages taken to select a appropriate alternate from those that are required for realizing a definite objective. The decision making process involves following steps:

(i) Determining the set alternates/substitutes
(ii) Evaluate the substitutes;
(iii) Compare between the substitutes

Semantic orientation for different words found and after aggregation the strength of opinions is negative and positive both. Aggregation can be done in different ways i.e. taking max, sum, prob. or custom function can also be used but here sum is used for aggregation. After calculations we get strength towards positivity is slightly greater than negativity. As overall negative orientation is 0.3626 and overall positive orientation is o.4252. So over all review about Samsung mobile phones is positive.

And in future try to make a FIS system which will interpret the reviews and find strength as accurate and clear as human brain. Also try to minimize the error in calculating the membership value by creating membership function that give result with minimum error.

VII. REFERENCES

[1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, Volumn 2, Issue 1-2 ppl-135, January 2008.

[2] Ritesh Srivastva, MPS Bhatia, H K Srivastava, "Exploiting Grammatical Dependencies forbFine-Grained Opinion Mining", IEEE International Conference on Computer & Communication 17-19 Septermber, 2010.

[3] . www.facebook.com.

[4] Fuzzy Sets By L.A. Zadeh, Information and Control 8, pp338-353 (1965).

[5] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning (III), Information Science, ol. 9 pp 43-80 (1975).

[6] Mamdani, E.H. and S. Assilian, "An experiment in linguistic Synthesis with a fuzzy logic controller, "International Journal of Man-Machine Studies, Vol. 7, No. 1, pp1-13, (1975).

[7] L.A. Zadeh, Fuzzy logic and approximate reasoning, Synthese, 1975, Vol.30, Issue 3-4, 407-428,1975.

[8] M. Hu and B. Liu, "Mining and summarizing customer reviews," proceedings of the ACM SIGKDD Conference on Knowledge Discovery abd Data Mining(KDD),pp.168-177,2004.

[9] Minqing Hu and Bing Liu "Mining opinion features in customer reviews" , Proceedings of the 19th international conference on Artificial intelligence, 2004.

[10] A. Esuli and F.Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining, "Proceedings of language Resources and Evaluation (LREC), 20006.

[11] Ritesh Srivastava and M.P.S Bhatia, "Quantifying Modified Opinion Strength: A Fuzzy Inference System for Sentiment Analysis", International Conference on Advance Computing, Communications and Informatics (ICACCI), 2013.