Volume 6, Issue 3, March 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Customer Feedback Evaluation System Using Feature Based Opinion Mining

[1]A. Kesavaram Kumar, [2]A. Periya Nayaki, [3]M. Indra Devi
[1] PG Student, [2] Assistant Professor, [3] Professor
[1, 2, 3] Department of Computer Science and Engineering, Kamaraj College of Engineering and Technology
Virudhunagar, Tamilnadu, India

***Abstract — In present competitive world the rapid growth of the web has been excitable increase in the user generated contents in the form of customer feedbacks, blogs, social networks, discussion forums, etc. Most of the contents are stored in the form of unstructured or semi structured data, where extracting of knowledge is a challenging task. Feedback is one of the most crucial parts of any educational institute. Feedbacks are collected and analyzed for various purposes like evaluation of products, manufacturer etc. for the better improvement of the company. Therefore evaluations and accomplishment monitoring of a product is also significant factor of the education system. This project proposes a customer feedback evaluation system for a particular product. This system uses opinion mining to evaluate the overall feedback scores. Opinion Mining is effective research fields that encompass natural language processing and text analysis techniques. Initially, this system collects customer feedback with the help of an online feedback form. The online feedback form consists of a free-form input text box. Every customer writes in detail about the performance of product in input text box. The text entered by the customer in input text box is then used for labeling the feedback as either positive or negative. The proposed methodology using ontology at the feature selection stage to progress feature based opinion mining. It helps to analyze and finding individual product quality performance and opinion of various customers about those product features. This analysis helps to identify feature-wise good and bad aspects of a particular product. The proposed approach provides a new SCA (Score Calculation Algorithm) based method for score calculation computed on various features from the customer opinions. The methodology has been implemented and tested in the canon digital camera feedback dataset. Evaluating result shows that the average accuracy of overall documents correctly classified as 81.11%.***

***Keywords- Opinion Mining, Text Classification, Feature Extraction***

## I. INTRODUCTION

Human's day to day life is filled with emotions and opinions. We cannot dream up the world without them. Emotions and opinions play a vital role in all human actions. They lead the human life by determine the way we think, what we do and how we act. Human beings require fast, accurate and concise information, so that they can make quick and genuine decisions. A person to be participating in discussions and express their points of view is quite natural. One's viewpoint about a subject can be either positive or negative, which is mentioned as the polarity of the opinion.

Although various methods and attitudes have been applied to opinion mining, two most important methodologies can be distinguished (i) the goal of opinion mining is to classify entire documents as having a positive or negative polarity (ii) approaches whose aim is to discriminate sentimental from non-sentimental sections.

In detail, classifying opinions at the document level or sentence level does not demonstrate what the user likes and dislikes about the entity. A positive document on an object does not specify that the user has affirmative opinions on all aspects of that object. In the same way, it is beyond the bounds of possibility to ensure that a negative document signifies that the user dislikes everything about the object. In a document (e.g., a customer review), the user typically writes about both the positive and negative aspects of the object[1]. To attain such comprehensive aspects, it is necessary to perform feature-based opinion mining in an effort to identify the features in the opinion and categorize the sentiments of the opinion for each of these features [4]. The majority of the researchers focus on the products and its services. In this paper, a Customer feedback evaluation system has to be designed. Regarding customer feedback evaluation, sentiment analysis is applied to categorize emotions and opinions about the performance of product. Sentiment analysis is enforced to classify the comment data into optimistic and pessimistic category to evaluate overall performance of product. Customers give feedback about different qualitative characteristic of product, which can be either optimistic or pessimistic. Here, Ontology based feature extraction and newly proposed score calculation method is used to evaluate customer feedback into positive and negative class and ranking their feature performances, based on customer reviews.

 

## II. LITERATURE SURVEY

In 2011, [13] V. K. Singh et al designed an automated web-based course feedback system that uses opinion mining to calculate the overall feedback scores. The system collect user response through an online form. They have used an unsupervised SO-PMI-IR based approach for opinion mining. They evaluated the system for more than 20 courses with a total of about 1000 reviews.

In 2013, [14] C.Sivagami et al proposed the domain ontology to extract the related generic category, in order to find the class of the movie based on the theme (Comedy, Action, Thriller, Tragedy, Horror, Sci-Fi, Family). They can viewed the vigour or failing of the movie more detail which is useful for further development and improvement of the movie.

In 2014, [8] Isidro Penalver et al proposed an innovative opinion mining methodology. The main goals of the proposed methodology are to improve feature-based opinion mining by using ontologies at the feature selection stage, and to provide a new vector analysis-based method for sentiment analysis. This methodology has been implemented and scrupulously tested in a real-world movie review-themed scenario, yielding very hopeful results when compared with other conventional approaches.

In 2014, [4] Balaji Jagtap et al proposed automate product feedback assessment system. Data collect from customer is larger in size to concluding result is tricky task. They were evaluate sentiment analysis with help of HMM and SVM base hybrid sentiment classification model with feature extraction method. Analysis helps to provide better understanding of customer opinion about product which also significant to improve the superiority of education. Hybrid approach works well for composite data.

In 2015, [16] Trisha Patel et al have collected feedback from parents regarding the overall functioning of the institute during various parents' meetings held at the institute. The reported feedback is extracted and transformed into *.doc files and supplied in tool for analysis. GATE is a tool narrating total positive and negative statements as a result of analysis. This research paper combines the data mining with natural language processing to extract the nuggets of knowledge from massive volumes of the customer feedback dataset of faculty performance.

## III. PROBLEM DESCRIPTION

The objective is to extracting the features of the customer reviews and their opinions. Opinions of different features from customer feedback are classified and summarized by a feature based opinion mining technique. This information will be more helpful to make decisions for better improvement of product quality.

The opinion mining dilemma can be etiquette defined as follows: Given a set of documents D, a sentiment classifier categorize each document $d \in D$ into one of the two classes, optimistic and pessimistic. optimistic means that *d* expresses a positive attitude and pessimistic means that *d* expresses a negative attitude.

Classical sentiment classification attempts to allocate the review documents either positive or negative class, but it fails to discover what the reviewer or opinion holder likes or dislikes. A feedback characterized ‘positive’ indicates that the customer who has written the feedback liked the performance attributes of particular product. On the other hand, a feedback characterized ‘negative’ signify that the customer who wrote the feedback did not have a superior experience with the performance attributes of the product. In an evaluative document (*Customer Feedback*), the opinion holder (*customer*) usually writes both positive and negative feature of the performance about product, although the general opinion on product performance may be positive or negative.

## IV. PROPOSED METHODOLOGY

The proposed work use domain ontology to progress the opinion analysis process. projected architecture consequences in a more detailed analysis of customer opinions in three proportions: (i) the common customer judgment (ii) features acknowledged in the customer opinion (iii) new SCA based method for feature score calculation. This proposed approach is not only to use of a single specific subject but for different subjects (e.g., Customer Review, Events, Product) at a time, since the proposed method is capable of analyzing and classifying the sentiments of each topic discussed by opinion holder (customer) in their opinion set.

We make utilizing of the SentiWordNet opinion word lexicon to facilitate identify the opinion- correlated words in the document. Our projected approach compact with feature classification by using domain ontology. In this paper, document-level opinion analysis obtain. Each customer opinion may broaden to an average of a hundred lines of text, so it is general for customer to provide feedback about the performance of product within the same opinion. In this position the sentiment analysis method must consequently meet a number of additional confront. The subsequent portion comprises the Module Descriptions, Result and Performance evaluation, Comparison.

## V. MODULE DESCRIPTION

The proposed work is to get better feature-based opinion mining by make use of ontology in the selection of features shown in Fig.1. The complete framework to achieve these goals is given in fig.1. There are five major modules namely,

1. Data Collection,
2. Data Pre-processing and Tagging,
3. Ontology based feature extraction,
4. Score Calculation,
5. Feature Visualize and Ranking

 

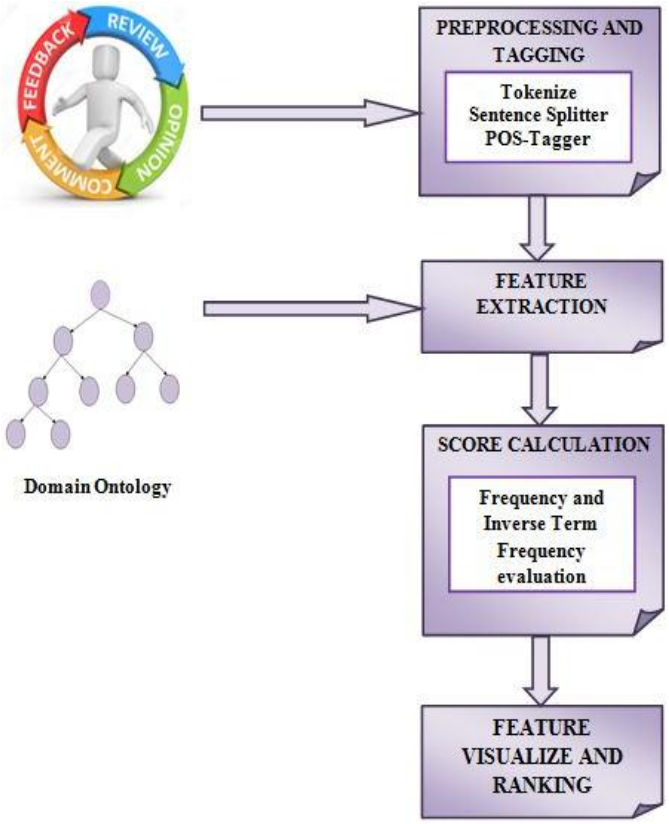


Fig. 1 Proposed System Architecture

### *A. Data Collection*

We have draft a web based interface to accept customer feedbacks. The online feedback form have been designed using HTML and Java Script. It contains a free form text box where a customer can write on the whole experience about the performance of particular product. Every customer write in detail about his/her experience about the product and their performance. After the customer completes the feedback form, the whole data entered is stored in the database server. The data entered by the customers are stored in a database designed which was implemented in mysql. The text entered in the database is then used for further estimation.

### *B. Data Pre-Processing and Tagging*

Data is collected from customer, which enclose noisy data and unnecessary text which does not contribute any kind of opinion. In addition, it may interrupt opinion analysis process. As a result, pre-processing the data is necessary before the sentiment classification. Pre-processing deals with Stemming, Tokenization task which will give correct data for classification of opinion.

In view of the fact that our intent is to extract performance features and the opinions from text documents, all pre-processed sentences are tagged using the Stanford Log-linear Part-Of-Speech Tagger framework (http://nlp.stanford.edu/software/tagger.shtml) which assigns Parts-Of-Speech (POS) tags to corresponding English words based on the situation in which they appear. A Part-Of-Speech Tagger (POS Tagger) is a software to reads text in a certain language and assigns parts of speech to each word (and other tokens), such as noun, verb, adjective, etc. The approach projected here is mainly designed to work with customer opinion texts written in English.

The POS information is used to situate different types of information in the text documents. In general, noun phrases correspond to performance features, adjectives represent opinions. For as much as, it is noticed that opinion words and product features are not individualistic of each other rather directly or indirectly inter-related through some semantic relations. Here is a sample sentence and the corresponding POS tagged sentence using Stanford Parser as shown in fig. 2(a) and 2(b) respectively.

Canon powershot g3 [+3] ##i recently purchased the canon powershot g3 and am extremely satisfied with the purchase.
use[+2]##the camera is very easy to use , in fact on a recent trip this past week i was asked to take a picture of a vacationing elderly group.

Fig. 2(a) Sample Sentence

 

canon/NN powershot/NN g/NN +/NN i/FW recently/RB purchased/VBD the/DT canon/NN powershot/NN g/NN and/CC am/VBP extremely/RB satisfied/VBN with/IN the/DT purchase/NN

use/NN +/VBD the/DT camera/NN is/VBZ very/RB easy/JJ to/TO use/VB in/IN fact/NN on/IN a/DT recent/JJ trip/NN this/DT past/JJ week/NN i/FW was/VBD asked/VBN to/TO take/VB a/DT picture/NN of/IN a/DT vacationing/VBG elderly/JJ group/NN

Fig. 2(b) POS tagged sentence

### *C. Ontology Based Feature Extraction*

Feature Extraction aspire to find what customers likes and dislikes about the performance of particular product. Therefore how to find out the performance features that customers talk about is an important step.

Ontology is used, in order to take out the features included in the customers opinions. This module collect a corpus of opinions as input. The sentences in the body of text that hold the classes, individuals, object properties and data types of the domain ontology are then identified from this input. Formerly the features have been recognized, they are grouped in accordance with their semantic distance. The features identified within the sentence shown in Fig.2 (a) are "canon power-shot g", "picture", "camera" etc. since the reviews, overall performance features of product are extract and stored in a database then its matching opinion words are extracted from these reviews.

### *D. Score Calculation*

Formerly the features in the customer opinions have been recognized, a score that signify the consequence of a given feature is calculated for each retrieved features. Since performance features are the key noun axiom on which opinions are applied, so a realistic collection of performance features is acknowledged using term frequency (tf) and inverse document frequency (idf).

Term Frequency, which deals how frequently a term take place in a document. Inverse Document Frequency, which measures how significant a term is in a document. The tf-idf value for each noun phrase is calculated using equations (1) and (2) where,

$$\textbf{TF(word)} = \frac{\textbf{Number of times word appears in a document}}{\textbf{Total number of terms in the document}} \qquad \textbf{(1)}$$

$$\textbf{IDF(word)} = \log_e\left(\frac{\textbf{Total number of documents}}{\textbf{Number of documents with word in it}}\right) \qquad \textbf{(2)}$$

After calculating the Term Frequency (TF) and Inverse Document Frequency (IDF) , values are stored in a database. The score value of the features in each customer's opinion is calculated by using newly proposed Score calculation method. The proposed SCA algorithm describes that the Score values can be deliberate by fetching the Term frequency value and Inverse Document frequency value that are already stored in a database and product of these two values with random value 'r' gives the score evaluation value. The random value 'r' has been taken for the reason that it has been found that in SentiWordNet that the difference in score values of the word. The score value of the characteristic in each customer estimation is calculated by proposed methodology is shown in equation (3) where,

$$\textbf{Score} = \textbf{TF (t) * IDF (t) *r} \qquad \textbf{(3)}$$

Thus the Score can be calculated and stored in database for further processing.

### *E. Feature Visualize and Ranking*

Once the score can be calculated, the score values of individual features of documents are stored in database. This module fetches the score values of each feature from an individual document stored in database. As a result, it summarize and display the features along with appropriate score values in the order of document based. Here we used the symbol "*" to express the score values. The score values can evaluated for 5 grades and the ranking of features can be determined as shown below.

**Score value = 1 * ---- 1 to 20 →Worst**
**Score value = 2 **---- 21 to 40→Bad**
**Score value = 3 ***---- 41 to 60→Fair**
**Score value = 4 **** ---- 61 to 80→Good**
**Score value = 5 *****---- 81to100→Excellent**

## VI. RESULT EVALUATION

In this portion, we present the tentative details of the projected opinion mining system. For experimental purposes, we have inserted a canon digital camera feedback dataset collected from the source amazon.com which consists of feature performance reviews about the canon digital camera. Our experiment comprises 300 opinions collected from

 

multiple numbers of customers. The general formula used to calculate accuracy of features correctly classified as given in equation (4),

$$\textbf{Accuracy}=(t/n) * 100 \qquad (4)$$

Where t- Total number of features correctly classified, n- Total number of samples (customers)

In our work, we assigned a grades for individual score values. Here we used the symbol "*" to express the score values. The score values can evaluated for 5 grades and the ranking of features can be determined as shown in above section. As on analysis the documents on each class, original (%) accuracy of individual document can be calculated by using equation 4 and the values are exhibit in Table I.

Table I. The sentimental analysis (%) accuracy of document

| Cannon Digital Camera Features | Result of Sentimental analysis positive opinions (%) | Result of Sentimental analysis negative opinions (%) |
|---|---|---|
| Four Megapixel | 60 | 40 |
| Metering Option | 80 | 20 |
| LCD | 80 | 20 |
| Software | 80 | 20 |
| Spot Metering | 60 | 40 |
| Control | 80 | 20 |
| Optimal Zoom | 80 | 20 |
| Digital Zoom | 60 | 40 |
| Picture | 80 | 20 |
| Auto Focus | 60 | 40 |
| Auto setting | 80 | 20 |
| Canon Power shot g | 100 | 0 |
| Camera | 80 | 20 |
| Feature | 80 | 20 |
| Feel | 80 | 20 |
| Speed | 60 | 40 |
| Photo | 100 | 0 |
| Manual | 60 | 40 |

The score value can be calculated based on our proposed system for all the input documents. The obtained accuracy (%) of correctly classified of document can be calculated by using equation (5) and the values are shown in Table II.

$$\textbf{Accuracy}= \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (5)$$

Table II. Accuracy (%) of features correctly classified based on SCA approach (rounded values)

| | Features | Correctly Classified % |
|---|---|---|
| Canon Digital Camera | Four Megapixels | 80 |
| | Metering Option | 80 |
| | LCD | 60 |
| | Software | 80 |
| | Spot Metering | 100 |
| | Control | 100 |
| | Optimal Zoom | 40 |
| | Digital Zoom | 100 |
| | Picture | 80 |
| | Auto Focus | 80 |
| | Auto setting | 100 |
| | Canon Power shot g | 80 |
| | Camera | 80 |
| | Feature | 100 |
| | Feel | 100 |
| | Speed | 60 |
| | Photo | 40 |
| | Manual | 100 |
| | Total Average | 81.11 |

As the result shown in TABLE II, the maximum correctly classified percentage is 100%. The minimum correctly classified percentage is 40%. Finally, the average success rate of canon digital camera dataset by the proposed approach is 81.11 % correctly classified.

 

## VII. COMPARISION

Comparison made between obtained percent of positive feature by proposed automated system and percentage of positive feature in original document. From the Fig. 3 it is clear that the average percentage of the positive opinion about the features four megapixels, metering options, LCD are 60%, 80%, 80% (values are rounded) in sentimental analysis of document. The proposed automated system obtained percentage of positive opinion about the features four megapixels, metering options, LCD are 80%, 60%, 40% (values are rounded). Thus the average accuracy of entire documents correctly classified as 81.11%.

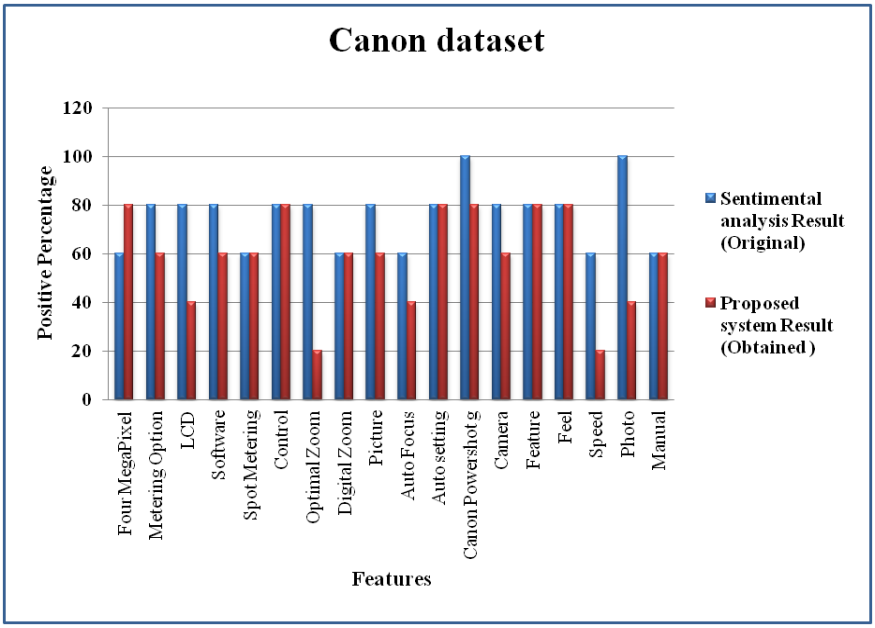


Fig. 3 Comparison of Proposed system

## VIII. CONCLUSION

Opinion mining is an promising research field of data mining used to extract the precious knowledge from huge volume of users remarks, reviews and feedback on any topic or product or event etc. We proposed an innovative methodology for evaluating customer feedback about the performance of the product by using Opinion Mining. Extraction of the product's performance feature is an significant task of review mining and summarization. This work tried to find out better product performance features from customer reviews. The proposed approach is based on three special stages: (i) feature identification, (ii) a new SCA based method for Score calculation computed on various features in the customer opinions (iii) summarize and display the features along with appropriate score values in the order of document based. Here firstly identified the features based on ontology based feature identification method. Secondly calculated the TF-IDF values of each features and evaluate score values for each features. Finally this system ranked the features on the origin of their score rates. The performance estimation result shows that the average accuracy of correctly classified feature was found to be 81.11%. This result indicates that the proposed techniques are effective in performing their tasks.

## ACKNOWLEDGMENT

A special thanks to Dr. M. Indra Devi (Professor) and Ms. A. Periya Nayaki (Asst.prof) Department of Computer Science and Engineering, for their valuable research advices and encouragement. A personal gratitude to my friends and my family for their inspirit reinforce.

## REFERENCES

[1] T.Ahmad, M.N. Doja (2012). Rule based system for enhancing recall for feature mining from short sentences in customer review documents. International Journal on Computer Science & Engineering, 4(6).

[2] Ayesha Rashid, Sana Asif, Naveed Anwer Butt, Imran Ashraf (2013). Feature Level Opinion Mining of Educational Customer Feedback Data using Sequential Pattern Mining and Association Rule Mining. International Journal of Computer Applications (0975 – 8887) Volume 81 – No.10.

[3] Ayesha Rashid, Naveed Anwer, Muddaser Iqbal, Muhammad Sher (2013). **"**A Survey Paper: Areas, Techniques and Challenges of Opinion Mining", Int.Journal of Computer Science Issues, Vol. 10, Issue 6, No 2.

[4] Balaji Jagtap, Virendrakumar Dhotre (2014). SVM and HMM Based Hybrid Approach of Sentiment Analysis for Product Feedback Assessment. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 3, 2278-6856.

[5] L. Chen, L. Qi and F. Wang (2012). Comparison of feature-level learning methods for mining online consumer reviews. Expert Systems with Applications, 39, 9588–9601.

[6] M. Eirinaki, S. Pisal and J. Singh (2012). Feature-based opinion mining and ranking. Journal of Computer and System Sciences, 78(4), 1175–1184.

 

[7] R. Feldman (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82–89.

[8] Isidro Peñalver-Martinez, Francisco Garcia-Sanchez, Rafael Valencia-Garcia, Miguel Ángel Rodríguez-García, Valentín Moreno, Anabel Fraga, Jose Luis Sánchez-Cervantes (2014). Feature-based opinion mining through ontologies. Expert Systems with Applications 41 (2014) 5995–6008, 0957-4174/@2014 Elsevier Ltd.

[9] E. Kontopoulos, C. Berberidis, T. Dergiades and N. Bassiliades (2013). Ontologybased sentiment analysis of twitter posts. Expert Systems with Applications, 40, 4065–4074.

[10] Larissa A. de Freitas, Renata Vieira (2013). Ontology-based Feature Level Opinion Mining for Portuguese Reviews. International World Wide Web Conference Committee (IW3C2).May 13–17, 2013, ACM 978-1-4503-2038-2/13/05.

[11] H. J. Min, & J. C Park (2012). Identifying helpful reviews based on customer's mentions about experiences. Expert Systems with Applications, 39(15), 11830–11838.

[12] Samad Hasan Basari, Abd., Burairah Hussin, Gede Pramudya Ananta, I., Junta Zeniarja (2013). Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization. Procedia Engineering Volume 53, Pages 453–462.

[13] V. K. Singh, P. Kumari, A. Singh, J. Thapa (2011). An Automated Course Feedback System using Opinion Mining. IEEE World Congress on Information and Communication Technologies, 978-1-4673-0126-8/11.

[14] C. Sivagami, S.C. Punitha (2013). Ontology Based Sentiment Clustering Of Movie Review. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 4, 2278-6856, (pp.375-378).

[15] M. Tarare, M. Manwani,A. Paidlewar,S. Maturkar, P. Chaudhari, J.V Shiral (2014). Feedback Management System for Evaluating And Generating Monthly Report. International Journal of Emerging Technology and Advanced Engineering,Volume 4, Issue 3, 2250-2459.

[16] Trisha Patel, Jaimin Undavia, Atul Patel (2015). Sentiment Analysis of Parents Feedback for Educational Institutes. International Journal of Innovative and Emerging Research in Engineering Volume 2, Issue 3, 2394-5494.

[17] Vibha Soni, Meenakshi R Patel (2014). Unsupervised Opinion Mining From Text Reviews Using SentiWordNet. International Journal of Computer Trends and Technology (IJCTT) – volume 11 number 5, 2231-2803.

[18] G. Vinodhini, L. Srisubha, R.M. Chandrasekaran(2012). Feature Based Opinion Mining For Customer Reviews. The International Journal of Information Technology and Engineering, ISSN: 2229-7367, (pp. 73-78).

[19] Z. Zhai,, H. Xu,B. Kang and P. Jia (2011). Exploiting effective features for Chinese sentiment classification. Expert Systems with Applications, 38, 9139–9146.

[20] L.Zhao and C. Li (2009). Ontology based opinion mining for movie reviews. In Proceedings of the 3rd international conference on knowledge science, engineering and management (pp.204–214).

 