

Lexicon based Sentiment Analysis of Parent Feedback to Evaluate their Satisfaction Level

Bhagyashree Gore

Student, P.G. Department of Computer Science,
S.N.D.T WU
Santa Cruz – Mumbai -49

Anita N. Chaware, PhD

Associate Professor,
P.G. Department of Computer Science,
S.N.D.T WU, Santa Cruz – Mumbai-49

ABSTRACT

It has become important to measure the quality of output of the teaching learning process. There is huge research going on improving the teaching learning process. Among the many one is to take the feedback of all the stakeholders of the education system. Along with the Students' feedback, Parents feedback is also one of the important evaluation factors while considering the evaluation of the teaching learning process of the Institute/ University and can be taken using any of the existing and effective channels of communication. This helps the parents get involved in decisions making of the institute / University, which further helps to build strong relationships between the parent and the university encouraging the involvement in the students learning and progress. In this paper we have collected the feedback in the textual format in paragraphs and is analyzed using lexicon based sentiment analysis approach to identify the parents positive or negative attitude.

The tool designed is collecting the parents textual feedback automatically and analyzes it using lexicon based approach to predict the level of Satisfaction of the parents w. r. to institute/university. A database of English sentiment words is created for reference in the domain of satisfaction level with the opinion score assigned to it. We extracted data from parent's comments and then analyze the level of positive and negative opinion. The opinion result of parents Satisfaction on teaching learning process is represented as to whether strongly positive, positive, strongly negative, negative, or neutral.

Keywords

Opinion mining; sentiment analysis; parents feedback; lexicon based; polarity;

1. INTRODUCTION

Parents, an alert stakeholders of education system, wish to have information about the curriculum, the subject areas that their wards are studying and the demands that will be placed on their children in the forthcoming year.

To enhance the perspective plan of institutional enhancement gathering feedback of all the stakeholders for the organization is very important aspect of that organization. From the collected feedback, measuring growth, preparing the reports in form of outcomes and visualizations, using this reports for decision making to expand the quality of their services has become mandatory job for the research team of any organization. This feedback is collected using measurement techniques such as surveys, academic assessment, and observation [1] and is required by the funding agencies for the review of these institutions /organization, all over the world. The same is applied to the education institutions for their prospects..

Parent satisfaction has been made an indispensable element of evaluations in education system [1]. The parent satisfaction data in form of questionnaire or textual is useful for several reasons[2]. First, parents know their children well and have the main role of guiding their child's development. Second, parent satisfaction or dissatisfaction can be used to improve the decision making of the institution. Finally, parents feedback can influence other community organizations as to the effectiveness of a program for funding agencies, etc...[3]

Once the data is collected from the parents, either in form of qualitative data or quantitative data, it is analyzed using any statistical packages.

For the textual data Sentiment analysis is used to analyze the feelings, opinions of person about a thing, a place, an object etc. It is also known as opinion mining, appraisal mining, subjectivity analysis etc. It is part of data mining in which, we extract the subjective data from large volume of data and then analyze that data for checking the patterns present in data. Sentiment Analysis also commonly known as Opinion Mining comes under the domain of Text Mining and Natural Language Processing for automatic extraction, classification and summarization of sentiments and emotions expressed.[4]

The parents give their feedback using textual comments. It is difficult to analyze textual data hence the sentiment analysis is used to mine and analyze the parents' comments.

The main elements used in sentiment analysis are object, features, opinion, opinion orientation etc. Object can be anything which exists in real world. For example mobile. Features are the attributes of an object. Suppose mobile is an object then touch screen, camera, memory are the characteristics of mobile. Feature extraction process is based on the textual input and generates the extracted features in the forms of lexicons which further is processed to find the weight of the lexicons or polarity of the lexicons and hence the sentence.

2. PROBLEM STATEMENT

Polarity of a sentence shows the contrary qualities, powers, tendencies, forms, opinions or ideas etc. or expresses the quantity of measurement of the condition of being positive, negative or neutral with respect to some objects. For example "good" defines the positive polarity of the sentence having the word "good". But the English languages have many complex features added to the sentences. Examples include "not good" indicates negative polarity, subjective terms like "fiction" in "science fiction" appears as neutral, neutral terms like "red" in "red flower" appears as neutral and there are intensifier and mitigation terms. Intensifiers increases the sentiments of its adjacent term (e.g., "very good"), here very adds more points to the positive polarity. The mitigators such as "few" reduces the sentiment or polarity of the sentences (e.g., "few support").[4]

Thus the problem statements for our research are as follows:

1. The problem is to classifying the polarity of a given text data coming as a document, sentence, or feature/aspect level.
2. To expressed the opinion coming in form of feedback and to check whether it is positive , negative, or neutral

3. RELATED WORK

Sentiment analysis and opinion mining being a very hot topic in the research activity, Be its finding market mining or opinion mining. M. Hu and B. Liu did a product feature based research where they tried to extract the product features and gave product based summary[5].

Various online and offline tools are also available for opinion extraction, sentiment analysis and opinion summarization. The researchers are trying to develop better and better algorithms for such tools. Ku, Liang and Chen [6] proposed algorithm for opinion extraction, opinion summarization and tracking the opinion which may be used for multiple languages.

Document level analysis is done by R. McDonald [7] with a finer level of classification. In his research, he used a structural model with different parameters for document level analysis. With the help of Viterbi's algorithm the models with different parameters or arguments was solved.

Bottom_up approach to identify opinions present was used by Virmani, Vikrant Malhotra, RidhiTyagi [8] for calculating collaborated opinion value.

Many researchers came up with the preparation on online dictionary word net, SentiWordNet[9] and many prepared their domain database compared their dictionary with the word net, SentiWordNet, ANEW, General Inquirer, LabMT or AFINN.[9].

LI Caiqiang and Ma Junming [10] proposed online education teacher evaluation model based on opinion mining. This model gets an overall evaluation of each teacher. They didn't indicate the words polarity strength in their work.

P. Kaewyong, A. Sukprasert, N. Salim and F.A. Phang. [11] proposed an application of language modeling for faculty comment evaluation. They applied sentiment analysis to determine the rating of positivity or negativity of the comments given by the computer science students in a Faculty Evaluation Form.

In this paper also, To get the polarity of words, SentiWordNet is utilized as a lexical source. To identify the polarity, the authors constructed the language model in which the probability of positivity, negativity or neutral occurrence is stored based on data that was fed beforehand. The objective of this paper is to evaluate the rating of institute performance from parents comments by using lexicon based approach.

4. SENTIMENT ANALYSIS APPROACHES

There are two main approaches of sentiment analysis one is machine learning and other one is lexicon based approach. In machine learning the machine automatically detects the patterns in given set of data and analyze the data to check its polarity. But, in case of lexicon based approach the sentiment word dictionary is used to analyze the feedback. The feedback is compared with the sentiment words present in sentiment word database and then the calculations are done. The two

main approaches as shown in fig.1 are further classified into different types as given in section 4.1 and 4.2:

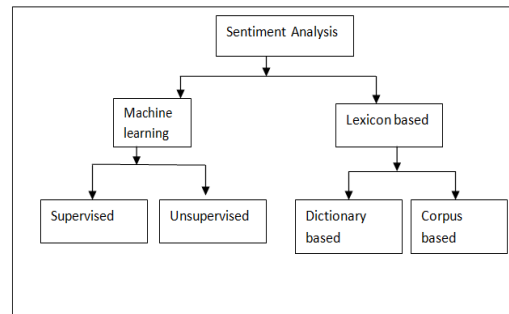


Fig.1: Sentiment Analysis Approaches modified form

4.1 Machine learning approach:

A. Supervised machine learning approach:

In this approach we use training data set. Using this training data the machine tries to learn the patterns and characteristics of an object and when new input is given to machine. Machine compares the given object with the training data and then determines the output. Suppose one we have training set having details of different fruits like shape, size and name of fruits. This is called supervised machine learning as it has supervisor which is trained using data which is already labeled i.e. data which is tagged with the correct answers. Example: When we enter shape and size of new fruit then machine gives the fruit name after comparing the given details with training data set. Suppose machine is given the shape of fruit as rounded shape with a depression at the top then classifier will compare that detail with already tagged data and identifies that fruit is an apple. Supervised machine learning is classified into two categories of algorithms: classification and regression.

B. Unsupervised machine learning approach:

In this approach the training data is not present. The classifier learns different patterns in data and then classifies the data according to some characteristics, features etc. Unsupervised machine learning is classified into two categories of algorithms: clustering and association. Clustering means analyzing all given data and then classifying the data into different groups with similar characteristics. K-means algorithm and aprori algorithm are the best examples of unsupervised machine learning. For example: Suppose different animals(cats and dogs) are given and the training set is not provided then machine will analyze their similarities and differences and then after identifying patterns it will classify those animals into different clusters.

4.2 Lexicon based approach:

In this approach we create a dictionary of words with opinion score assigned to it. There are two types of this approach:

A. Dictionary based approach:

In dictionary based approach, a set of words is taken as seed then these words are compared with standard dictionary which is available online. Then antonyms and synonyms are added to the set of words. This is iterative process. The process is stopped when no such antonyms and synonyms are found in dictionary. Dictionary based approach is more effective than corpus based approach.

B. Corpus based approach:

Corpus based approach is complex than dictionary based approach. In corpus based approach we create dictionary with context based words only. Suppose in this case we will create dictionary having words related to teaching. The score and polarity of the words are determined as per the domain.

For example: the word fast has different orientation in different scenarios.

- 1) The delivery of packages is very fast
- 2) Teacher talks very fast, hence it becomes difficult to understand things.

Consider above statements, the first statement is used in case of logistic companies where the word fast has positive orientation but in second statement the word fast is depicting negative feedback. Hence the polarity of words changes according to domains. The following table gives the major difference between Corpus based and dictionary base sentiment analysis.

4.3 Comparison between corpus based and dictionary based sentiment analysis approach:

Table 1: Showing the difference between corpus based and dictionary based sentiment analysis approach

Corpus based sentiment analysis	Dictionary based sentiment analysis
It suggests data driven approach where you have access to not only sentiment labels but to a context	It suggests you'd implement judging of sentiment based on presence of sentiment words(and perhaps some shorter context)
It is little complex than dictionary	This is referred to as the simplest one
It finds context dependent opinions	It does not give context dependent opinions

5. SENTIMENT WORD DATABASE CONSTRUCTION

This section presents the construction of sentiment word database. The lexicon can be created either manually [12] or expanding automatically from a seed of words. In many works lexicon the synonym is considered as a positive word and antonym is considered as a negative word [13]. In our work, the score of opinion words are defined by us in order to maintain the simplicity as our main focus was to create a database of words in the domain of satisfaction level.

The sentiment word database has different positive, negative words. The database also has blind negation words, negation words, adverbs, verb and noun etc. It has intensifiers. Intensifiers are of two type i.e. amplifiers and down toners. As the name suggests the amplifiers increase intensity of words and the down toner decreases the intensity of word. When word is preceded with the down toner the sentence is considered to be negative and vice versa.

Example-very good, in this case the word very is an intensifier which increases the intensity or effect of word

good. The remaining words are neutral. Table 1 shows the sample sentiment word dictionary used in this research work.

Table 2: Sentiment word database-snapshot

Words	Score
Achieve	2
Amuse	1
Comfort	1.5
Delight	2
Enjoy	2
Gratification	1.5
Happy	2
Unhappy	-2
Pleasure	2
Discontent	-1
Displeasure	-2
Dissatisfaction	-2
Upset	-0.25
Sad	-1
Need	-2
Poor	-1.5
Good	2

We have used internet for searching the words with sentiment score assigned between -2 to +2. The data base is limited to a number of positively and negatively tagged words. The total score is initialized to zero. Then it is checked if the lexicon is present in the dictionary of words and if present whether it is positive or negative and the score is updated accordingly. The final score will thus classify the text as positive, negative or neutral.

6. PROPOSED SYSTEM PROCESS:

The proposed system has some important steps. Those steps are used for conducting sentiment analysis and the steps as below

6.1)Data collection: This is very first step of sentiment analysis. The data is collected from parents . Parents gave feedback using google form. For the research purpose, only 50 parents data was analysed. The data collected thus from comments is saved in excel.

6.2) Data preprocessing: This is very important step. During this step of sentiment analysis, following steps are performed:

a)Tokenization: Tokenization is the process of dividing or breaking down the sentences into words, phrases.

For example: The school is good. Then this sentence will be divided into tokens as :



b) Lowercasing: the feedback data is converted into lowercase since the words present in sentiment word database are

present in lowercase. It becomes easy to compare the feedback data with the sentiment word database.

c) Removing irrelevant data: every language has some stop words. Stop words are those words which are not assigned any polarity score, usually the non-semantic words like articles, prepositions, conjunctions and pronouns. Articles such as ‘a’ and ‘the’ and pronouns such as ‘he,’ ‘they,’ and ‘I’ provide little or no information about sentiment. These words are removed from the sentences and the feedback data is cleaned. Consider above example: she is very nice person. In this case she and is are stop words which are not assigned score. Hence these words are removed from the sentence and array of sentiment word is created for every feedback.

Fig. 2 shows the stop words present in English language which were collected R studio using stopwords(‘english’) function just to list the stop words which are to be removed from the feedback before starting the processing. The screenshot below in fig.3. has the code required to remove the stop words those are not contributing to any kind of calculations.

"i"	"me"	"my"	"myself"	"we"
"our"	"ours"	"ourselves"	"you"	"your"
"yours"	"yourself"	"yourselves"	"he"	"him"
"his"	"himself"	"she"	"her"	"hers"
"herself"	"it"	"its"	"itself"	"they"
"them"	"their"	"theirs"	"themselves"	"what"
"which"	"who"	"whom"	"this"	"that"
"these"	"those"	"am"	"is"	"are"
"was"	"were"	"be"	"been"	"being"
"have"	"has"	"had"	"having"	"do"
"does"	"did"	"doing"	"would"	"should"
"could"	"ought"	"i'm"	"you're"	"he's"
"she's"	"it's"	"we're"	"they're"	"i've"
"you've"	"we've"	"they've"	"i'd"	"you'd"
"he'd"	"she'd"	"we'd"	"they'd"	"i'll"
"you'll"	"he'll"	"she'll"	"we'll"	"they'll"
"isn't"	"aren't"	"wasn't"	"weren't"	"hasn't"
"haven't"	"hadn't"	"doesn't"	"don't"	"didn't"
"won't"	"wouldn't"	"shan't"	"shouldn't"	"can't"
"cannot"	"couldn't"	"mustn't"	"let's"	"that's"
"who's"	"what's"	"here's"	"there's"	"when's"
"where's"	"why's"	"how's"	"a"	"an"

Fig.2: Stop words in English

```

1 using System;
2 using System.Collections.Generic;
3 using System.Linq;
4 using System.Text;
5 using System.Threading.Tasks;
6
7 namespace SentimentAnalysis
8 {
9     /// <summary>
10    /// Tool to remove unwanted words such as 'the' or 'a'.
11    /// </summary>
12    public class StopwordTool
13    {
14        /// <summary>
15        /// Words we want to remove.
16        /// </summary>
17        static Dictionary<string, bool> _stops = new Dictionary<string, bool>
18        {
19            { "a", true },
20            { "about", true },
21            { "above", true },
22            { "across", true },
23            { "after", true },
24            { "afterwards", true },
25            { "again", true },

```

Fig.3: Stop words removal code

The next step is opinion word identification. This system compares each tokenized word in the comment with blind negation word or negation word or positive opinion word or negative opinion word or intensifier word by using sentiment word database. And then the polarity scores are assigned to each word by using sentiment word database. The presence of the blind negation word indicates negative sentiment value eq-2.

In the following equations, ‘Ws’ is the semantic orientation score of combining words. ‘Sinf’ is the

intensifier value of word based on 100%. ‘Os’ is the score of opinion word from sentiment word database.

$$Ws=Os \quad \text{eq-1}$$

$$Ws=(100\%+Sinf) * Os \quad \text{eq-2}$$

$$Ws=(100\%+Sinf) * (100\%+Sinf) * Os \quad \text{eq-3}$$

$$Ws=Ws * (-1) \quad \text{eq-4}$$

If there is only one opinion word in a sentence, the

corresponding positive scores or negative scores are assigned using eq-1. If one intensifier word and one opinion word are found together, i.e. the location of intensifier word is adjacent with the location of opinion word, eq-2 is used to get the semantic orientation score of combining words. If two intensifier words and one opinion word are found in a sentence, moreover the index of first intensifier word must be the index by reducing two of the index of opinion word, eq-3 is used. If a negation word in front of the opinion word is found in a sentence, reversed polarity scores are given by eq-4.

The semantic orientation score of combining words in all sentences are summed up to get the total polarity scores by eq-5. In eq-5, ‘PTs’ is .the total polarity score of all words in all sentences from one comment. ‘Wsi’ is the semantic orientation score of combining words for ith term in one comment. ‘i’ is the order of combining opinion words from 1 to n. ‘n’ is the total number of combining opinion words in all sentences from one comment. ‘T’ is set of teaching sentiment terms from sentiment word database. ‘PTsi’ is the total polarity score of ith term for all

comments. ‘N’ is the total number of opinion words in all comments. ‘P’ is the average polarity scores of all comments.

$$\text{Total} = \sum(i=1 \text{ to } n) Ws \quad \text{eq-5}$$

$$\text{Average} = \text{Total}/N \quad \text{eq-6}$$

Where, N is count of (opinion) sentiment words from comments

Using this formula we calculate the sum of the opinion score assigned to the sentiment words which are obtained from the parents’ comments about college and teachers. The opinion score is collected from the sentiment word database. After calculating sum of opinion score we take the sentiment word count from the parent comments. Then using this sum and count of opinion word we calculate the average opinion score. Depending upon this score we decide whether the feedback given by parents is positive, negative or neutral. The average polarity score of all comments can be calculated by eq-6.

If P>0 or <=1, the opinion result is defined as weakly positive.

If P>1 or <=2, the opinion result is defined as moderately positive.

If P>2, the opinion result is defined as strongly positive.

If P<0 or >=-1, the opinion result is defined as weakly negative.

If P<-1 or >=-2, the opinion result is defined as moderately negative.

If P<-2, the opinion result is defined as strongly negative.

If P=0, the opinion result is defined as neutral.

The whole process of lexicon based sentiment analysis used in this research is depicted in fig.4.

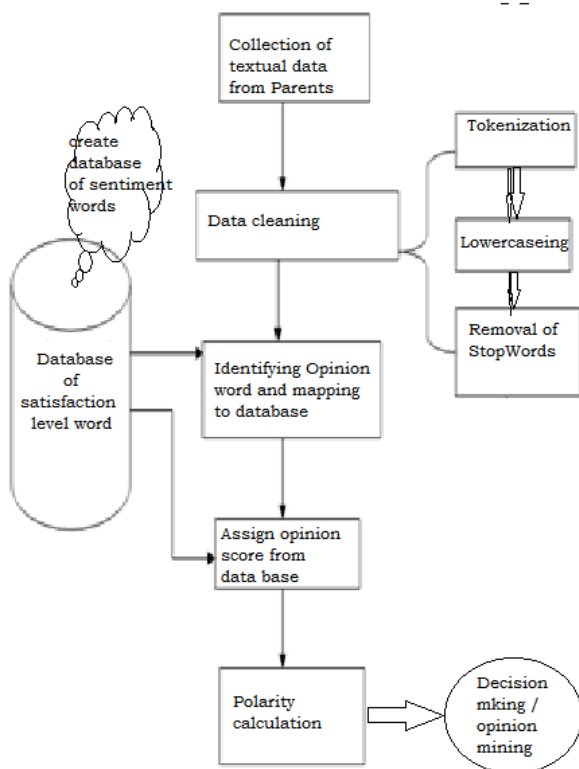


Fig.4: Flow of sentiment analysis explaining the whole process of Lexicon based Semantic analysis.

7. CASE STUDY AND RESULTS:

The opinion result of all comments is classified as strongly negative, or moderately negative, or weakly negative, or strongly positive, or moderately positive, or weakly positive or neutral.

Comment 1: The School has very good project oriented curriculum

There is only one opinion word i.e. good.

The opinion score calculation of comment 1 is very interesting $(3 \times [100\% + 50\%]) = 4.5$ by eq-2.

The total opinion scores in comment 1 is 4.5 by eq-5. Then we calculate average by eq-6 i.e. $4.5/1 = 4.5$.

The average is greater than 2 hence the feedback is Strongly positive.

Comment 2: School has no good teacher and teaching methods are not effective.

The opinion score calculation of comment 2 is not good $-3 = -3$, not effective $-2 = -2$ by eq-4.

There are two opinion words good and effective.

The total opinion scores in comment 2 is (-5) by eq-5. Then we calculate average by eq-6 $-5/2 = (-2.5)$.

The average is less than -2 hence the feedback is Strongly negative.

Comment 3: Principal is good teacher but her teaching techniques are not good.

The opinion score calculation of comment 3 is good $= 3$ by eq-1, not good $= -3$ by eq-4 and divided by total number of opinion words in all comments.

The total opinion scores in comment 2 is $(-3+3)=0$ by eq-5. Hence the feedback is neutral.

The below graph in fig.5 shows the all Parents Satisfaction level in relation to overall effectiveness of institute / University collected in the research of 50 parents. Fig.6 shows the calculation of the opinion score for all the 50 parents feedback.

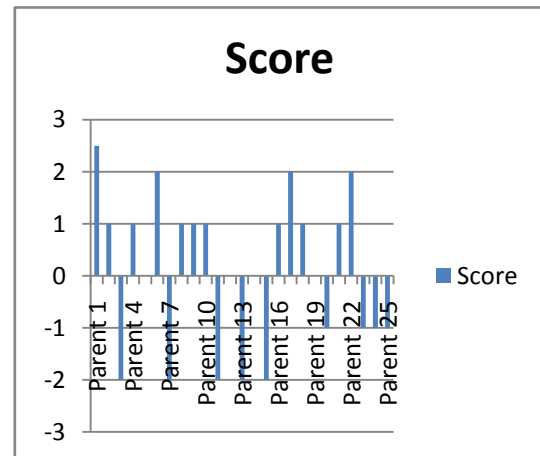


Fig.5: Graph showing Parents Satisfaction level

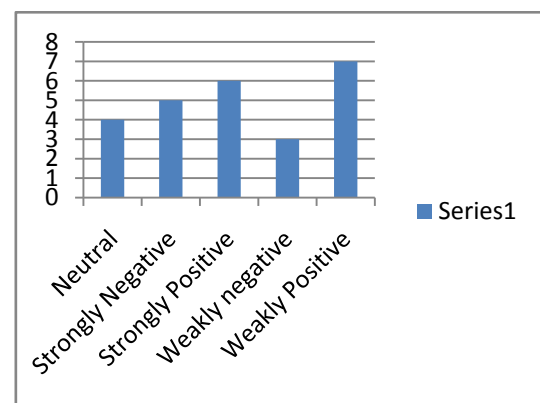


Fig.6: Total satisfaction polarity graph

The next job we did is to prepare an app for the Institute / University and store the feedback into the database and then summarize its polarity. The above screenshots, we have added only opinion words with the intensifiers and calculation are conducted by the system. The output is stored in the database for further reference.

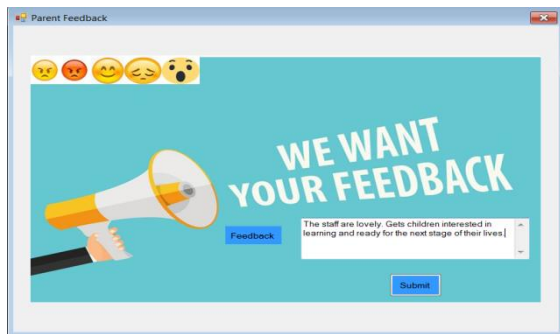


Fig.8: Screenshot of parent feedback tool

8. CONCLUSION

In this paper, A pilot study of lexicon based sentiment analysis is used to evaluate the level of teaching performance from parents' textual feedback comment. A database of English sentiment words is constructed to identify the polarity of words as a lexical source. The sentiment word database used in this research contains the opinion words concerning with the satisfaction level / domain to achieve the better result. Every opinion word in the database is given a value. The sentiment value is ranged from -2 to +2. This method analyze automatically the parents' feedback comments to strongly negative, or weakly negative, or strongly positive, or weakly positive or neutral category using the lexicons. The level of opinion result for the institute overall performance is given out from parents' feedback comments.

9. REFERENCES

- [1] Bell& Carrillo, "Characteristics of effective summer learning programs in practice." *New Directions for Youth Development*, 2007., doi:10.1002 /yd.212
- [2] Ratzon, N.Z., Zabaneh-Tannas, K., Ben-Hamp, L., & Bart, "Efficiency of the home parental programme in visual-motor home activity among first-grade children.", *Child: care, health, and development*, doi:10.1111/j.1365-2214.2009.1011.x
- [3] Feuerstein, A., "School characteristics and parent involvement: Influences on participation in children's schools". *Journal of Educational Research*, 94, (1), 29-40.
- [4] Luke Kien-Weng Tan¹ , Jin-Cheon Na¹ , Yin-Leng Theng¹ , and Kuiyu Chang², "Sentence-Level Sentiment Polarity Classification Using a Linguistic Approach", *ICADL*, 2011
- [5] M. Hu and B. Liu, "Mining and summarizing customer remarks," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 168–177, 2004.
- [6] Lun-Wei Ku, Yu-Ting Liang and Hsin-Hsi Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora", 2006 American Association for Artificial Intelligence.
- [7] R. McDonald , K. Hannan, T. Neylon, M. Wells, and J. Teynar, "Structured models for fine-to-coarse sentiment analysis," *Proceedings of the Association for Computational Linguistics (ACL)*, June 2007, 432-439.
- [8] D. Virmani, V.Malhotra, R. Tyagi, " Sentiment Analysis Using Collaborated Opinion Mining," *BhagwanParshuram Institute of Technology, New Delhi, India*, January, 2014.
- [9] V. Soni, M.R.Patel, "Unsupervised Opinion Mining From Text Reviews Using SentiWordNet," *International Journal of Computer Trends and Technology (IJCTT)*, vol 11, May 2014.
- [10] L. Caiqiang and M. Junming, "Research on Online Education Teacher Evaluation Model based on Opinion Mining," *National Conference on Information Technology and Computer Science (CITCS)*, 2012.
- [11] P. Kaewyong, A. Sukprasert, N. Salim and F.A. Phang, "The possibility of students' comments automatic interpret using Lexicon based sentiment analysis to teacher evaluation," *3rd International Conference on Artificial*
- [12] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexiconbased methods for sentiment analysis," *Association for Computational Linguistics*, vol. 37, Issue 2, pp. 267-307, June 2011.
- [13] Sujata Rani and Parteek Kumar, Thapar University, A Sentiment Analysis System to Improve Teaching and Learning
- [14] G. Angulakshmi, Dr. R. ManickaChezian, "An Analysis on Opinion Mining: Techniques and Tools," *International Journal of Advanced Research in Computer and Communication Engineering*, vol3, July 2014.
- [15] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, K.Miller, "WordNet: an on-line lexical database," *Oxford Univ. Press*, 1993.
- [16] Lun-Wei Ku, Yu-Ting Liang and Hsin-Hsi Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora," *American Association for Artificial Intelligence*, 2006.
- [17] Amitava Das et al , 2008, Topic-Based Bengali Opinion Summarization, *Coling 2008: Poster Volume*, pages 232–240, Beijing, August 2010.
- [18]Ahmed Abbasi, etl. "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums" *ACM Transactions on Information Systems*, Vol. 26, No. 3, Article 12, 2008.