



## ASPECT BASED SENTIMENT ANALYSIS OF HINDI TEXT REVIEW

Komal Garg

M.Tech Scholar: Dept. of Computer Science  
Sant Longowal Institute Of Engineering & Technology  
Sangrur (Punjab), 148106, India

Preetpal Kaur Buttar

Assistant Professor: Dept. of Computer Science  
Sant Longowal Institute Of Engineering & Technology  
Sangrur (Punjab), 148106, India

**Abstract:** Sentiment analysis (SA) is one of the fastest growing research areas in Natural Language Processing, making it challenging to keep track of all the activities in the area. Increase in user-generated content (UGC) has provided an important aspect for the researchers, industries and government(s) to mine this information. SA mine information from UGC on the basis of polarity as positive, negative or neutral. The problem domain, to which this research is concerned, is to find the sentiment and its respective aspect in the sentence and finally to calculate the overall sentiment score of entered Hindi text to classify each sentence as positive, negative and neutral. In this thesis, we work on the sentiment analysis by devolving an algorithm that identifying the sentiment according to proposed rules based on positions of conjunction, negation and aspects (nouns).

**Keywords:** Sentiment Analysis, Hindi Aspect, Conjunction, Negation

## I. INTRODUCTION

With the globalization of internet, the users and contributors to the increasing web space are rising tremendously. This huge amount of data generated everyday has introduced various challenges and opportunities in the research communities. The opinions of others have a critical impact in our daily decision-making processes. Now, in the era of internet, it is easy to collect diverse opinions from different all around the world. People look after the review sites such as, Amazon, eBay, Facebook, Twitter in order to get feedback of a product or service. In order to get an unbiased opinion one has to extract and read all the reviews which is not an easy task to perform. Sentiment Analysis [1] is the computational study of opinions, sentiments, and emotions expressed in text. It classifies the text written by user by predicting its polarity as either positive or negative. Finding the polarity of user review with respect to some feature (aspect) is known as aspect based sentiment analysis (ABSA). The term ‘aspect’ refers to an attribute of the entity (product) that has been looked after in a review.

*For example: I bought Activa Honda scooter a few days ago. It is a beautiful scooter and has a comfortable seat. The design of scooter is also attractive.*

The above review is analyzed and extracted views are represented in fig 1.

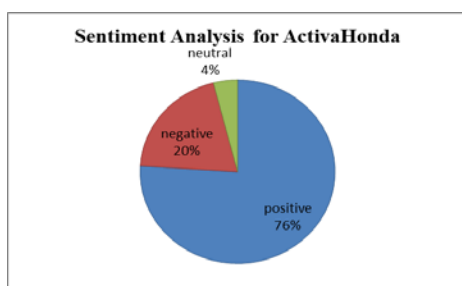


Figure 1

Our work concentrates on analyzing sentiments and its aspects for Hindi language. The main focus lies on understanding the issues while working on Hindi language and various approaches followed while performing sentiment classification for the user-generated content (UGC). The work done comprises of resource generation which involves building of datasets. We propose an algorithm to extract the aspect and the sentiment from the review (UGC). We try to handle two different cases while performing sentiment classification-

- 1) **Conjunction** - A conjunction is a part of speech that is used to connect words, phrases, clauses, or sentences. For Example: And, Yet, But, For, So, Whether, etc. Here are some conjunction used in Hindi: लेकिन, पर, मगर, एवं, और, etc
  - AND conjunction: polarity of the sentence before and after “and” conjunction must be same.
  - BUT conjunction: polarity of the sentence before and after “but” conjunction is different.
- 2) **Negation**- A negation is a refusal or denial of something. In Hindi, we learnt the word “*नहीं*” (naheen) for negation word “no”. It reverses the meaning of words when used with them [2]. For example-“I am happy” into its opposite denial “I am not happy”. Thus, changes the polarity of sentence.

## II. RELATED WORK

Various works on Sentiment Analysis have been done for the English language. Yao and Chen [3] have applied sentiment analysis and machine learning methods to study the relationship between the online reviews for a movie and the movie’s box office revenue performance. Ming and Ying have used support vector machine (SVM) as a sentiment polarity classifier [4]. A survey done by Pang B covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems [5]. Singh and Piryani have stated a new kind of domain specific feature-based heuristic for aspect-level sentiment analysis of movie reviews [6]. They have devised an aspect oriented scheme that analyses the

textual reviews of a movie and assign it a sentiment label on each aspect.

Sentiment analysis in Indian languages (especially Hindi) are still largely unexplored due to various challenges like non-availability of various resources and tools such as annotated corpora, lexicons, Part-of-Speech (PoS) tagger etc. Akhtar, Ekbal and Bhattacharyya created a dataset and build machine learning models for SA in order to demonstrate the powerful use of the dataset, and finally make the resource available to the community [7]. Joshi, A R, & Bhattacharyya, 2010 developed a fall-back methodology to do supposition examination for Hindi reports [8]. Balamurali et al. 2012 [9]; Balamurali et al. 2011 [10]; Bakliwal et al. 2012 [11]; Mittal et al. 2013 [12]; Sharma et al. 2014 [13]; Das and Bandyopadhyay, 2010a [14]; discuss the problems of sentiment analysis at the coarse-grained level with the aims to classify sentiments either at the sentence or document level in Indian languages.

### III. SENTIMENT CLASSIFICATION TECHNIQUES

The Sentiment Classification (SC) techniques, as indicated in Fig2, are examined [15]. Sentiment Classification procedures can be generally isolated into machine learning methodology, lexicon based approach and hybrid approach. Machine Learning (ML) algorithms are classified into supervised and unsupervised learning. ML algorithms are applicable to SA mostly belongs to supervised classification.

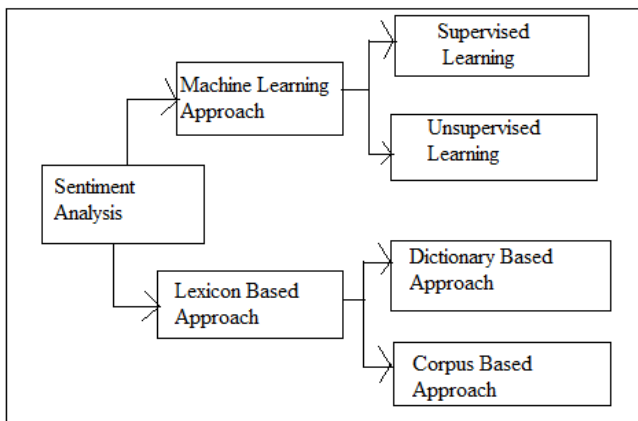


Figure 2

The Lexicon-based Approach depends on a sentiment lexicon, a gathering of known and precompiled opinion terms. It is divided into dictionary based approach and corpus-based approach which utilize statistical or semantic methods to discover sentiment polarity. It is the process of determining the correct meaning of an individual word. There are two strategies in this methodology. The dictionary based approach depends with respect to discovering supposition seed words, and afterward looks the lexicon of their equivalent words and antonyms. The corpus-based approach starts with a seed list of supposition words, and afterward finds other sentiment words in an expansive corpus to help in discovering conclusion words with connection particular introductions. This should be possible by utilizing factual or semantic systems.

To collect the sentiment word list, two approaches have been investigated: dictionary-based approach and corpus-based approach.

#### Dictionary-based approach

A small set of opinion words is collected manually with known orientations. Then, this set is grown by searching in the well-known corpora Word Net [16] or for their synonyms and antonyms. The newly found words are added to the seed list then the next iteration starts. The iterative process stops when no new words are found. After the process is completed, manual inspection can be carried out to remove or correct errors. [17] has presented the strategy of dictionary-based approach.

In [18], dictionary-based approach has been used to identify sentiment sentences in contextual advertising. An advertising strategy has been proposed for the improvement of ad relevance and user experience. Qui and He used syntactic parsing and sentiment dictionary and proposed a rule based approach to handle word extraction and consumers attitude identification in advertising keyword extraction.

### IV. RESEARCH GAP

A large number of broadcasting industries work on grasping person’s liking and disliking from their reviews. There are 100 plus million Hindi language speakers spread across the globe. The consideration region of Hindi language is growing across the web. Considering the study or research on sentiment Analysis, the Hindi language does not holds ample work. In today’s digital era, text is the best way of representing and conveying data or information, and our lives are soaked with textual data. Therefore there is a growing necessity to develop technology for Hindi language that assist us control and make understanding of the resulting information.

On real-life applications, to give a fully automated result or output is the ultimate aspired aim of all the sentiment analysis (research). A brilliant system should be smart enough to sum all the dispersed sentimental data (information) from the different blogs, news article and from reviews that are written and give us results. The part of any automatic system is to reduce user’s attempt and generate a good sensible output. There is no such application available that classifies the reviews as positive, negative and neutral for Hindi language. There is a need to analyze the Hindi language content and obtain judgment of opinions conveyed by people.

### V. PROBLEM DEFINITION

Many real-life applications require a detailed analysis as the user wants to know what opinions have been expressed on [19] [20] and for which object. For example, from the review of a product, a person (buyer) wants to know what features of the product have been praised and criticized by other consumers. For Example: “I bought a phone a few days ago. It is a nice phone. The voice quality is also clear. But, the camera quality is poor.”

Entity	Aspect	Sentiment
Phone	-	Positive
Phone	Camera Quality	Positive
Phone	Voice Quality	Negative

In this example, the phone and its voice quality expresses positive opinion but camera quality expresses negative opinion. Another challenge is to find out to what extent the sentence is positive or negative, i.e., what is the score of sentence. For Example: “It is a good story.”

Sentiment Word	Sentiment Score	Polarity
Good	0.3125	Positive

The problem domain, to which this research is concerned, is to find the sentiment and its respective aspect in the sentence and finally to calculate the overall sentiment score of entered Hindi text to classify each sentence as positive, negative and neutral. Our algorithm works according to the sentiment words and their respective sentiment scores stored in the database dictionary.

## VI. PROPOSED SYSTEM

We proposed basic aspect-level sentiment analysis system to analyze the aspects and sentiments in the given sentence and then classify the sentiment as positive, negative or neutral. Our approach used is dictionary-based in which we extract the aspects and sentiments given in a sentence; and calculate the sentiment score of sentence from database dictionary. Sentiment score is taken from SentiWordNet database for Hindi.

In this work, our focus is to handle two different cases:

- 1) Negation Handling - If difference between the position of negation word and sentiment word is 1, then polarity of the sentiment gets changed. *For Example: फिल्म अच्छी नहीं है* / Here, the review has negative polarity.
- 2) Conjunction Handling - If the review sentence contains conjunction words, two possible situations take place.
  - When conjunction word is “and”: polarity before the conjunction word and after the conjunction word must be same. If polarity of the sentence before और is positive and after और is negative, then reverse the polarity of sentence after conjunctive word और and vice versa. *For Example: फल मीठा है और स्वादिष्ट भी* / Here, polarity before और and after और is positive.
  - When conjunction word is “but”: polarity before the conjunction word and after the conjunction word must be opposite. If polarity on both sides of conjunction is same, then reverse the polarity of sentence after conjunction word लेकिन. *For Example: फोन अच्छा है लेकिन बैटरी बैकअप कम है* | Here, review before लेकिन is positive but review after लेकिन is negative.

## VII. METHODOLOGY USED

Implementation of algorithm is discussed as:-

- 1) **Review collection and storage-** Data generated by different people is available on web in various forms. Content that has been created and put out there by unpaid contributors is known as User Generated Content (UGC). It allow users to collaborate in a way that helps them be creative, connect with various brands and one another- enhancing the respective sites with their activities. The goal of the system is to enter the reviews that contain people’s opinions and store them in a database.
  - i. **Online News-** The online-news taking into account every day occasions, exercises and other data over the globe with a fast reachability and that too in a multilingual manner.
  - ii. **Weblogs-** Accumulation of web journals, arrangements of sites mostly as hyperlinks to be utilized as an instrument for proposal and recommendation to different web journals and sites.
  - iii. **Review Sites-** Sentiments are the choice makes for any client in making a buy. The client created audits for products and services are accessible on web. The sentiment classification user reviewer's information gathered from the sites like www.amazon.com which has a large number of item audits by consumers [21].
  - iv. **Facebook-** Facebook is a remunerating stage for marketing. Brands offer customers a pleasing experience. The Facebook Review empowers customers to rate check pages. Brands can determinedly review this experience and can shut it down if they believe they are getting suffocated in negative reviews.
- 2) **Creation of SentiWordNet Database-** Database of Hindi searching words is created and stored in database array. It contains the status of the array which tells whether the word is an aspect (noun) or a sentiment. It also contains the sentiment score, i.e., polarity of the database.
- 3) **Sentiment Classification-**
  - 3.1 **Parsing of review:** The input review is parsed into tokens and stored in string array [22].
  - 3.2 **Text extraction:** In this step each word is extracted from the string array and compared with each word of the database [22].
  - 3.3 **Sentiment classification:** If comparison is true, check the status of selected word whether it is an aspect or a sentiment. Extract its sentiment score from the database and save in variable (var.) result. If the selected word is an aspect, check the status of variables: position\_noun and position\_property\_noun which is already set to -1. If position\_noun = -1; then the position of selected word is passed to var. position\_noun, i.e., position\_noun = position of selected word. Else If position\_noun != -1; then the position of selected word is passed to var. position\_property\_noun, i.e., position\_property\_noun = position of selected word. If the selected word is a sentiment: update the var. position\_sentiment which is already set to 0, i.e., position\_sentiment= position of

selected word. Score of selected sentiment is updated to final answer, i.e., ans = result.

While performing sentiment classification, we try to handle two different cases. These are:

#### Case 1: Negation Handling

On traversing the array, if difference between the position of negation word and sentiment word is 1, then the polarity of var. result gets reversed, otherwise no change in result. If more than one sentiment word is found, final answer is updated as: ans1= ans1 + result.

#### Case 2: Conjunction Handling

On traversing the array, conjunction word “and”/ “but” is found, and position of sentiment word is less than position of conjunctive word and position of sentiment word is less than position of negation word, then: reverse the polarity of answer.

If sentiment word is present before conjunction “and” and one sentiment word is present after it: extract the score of first part of sentence (before conjunction) and save in var result1, then extract the score of second part of sentence (after conjunction) and save in var result2. If result1 is less than zero and result2 is greater than zero, then reverse the polarity of result2 and vice-versa.

If sentiment word is present before conjunction “but” and one sentiment word is present after it: extract the score of first part of sentence (before conjunction) and save in result1, then extract the score of second part of sentence (after conjunction) and save in result2. If result on both sides is same then change the polarity of result2. Final answer will be updated by adding both result1 and result2.

**4. Retrieval:** The final answer, i.e., ans1 is retrieved.

**5. Result Identification:** Find out the impact of review whether it is positive or negative.

- If result>0, review is positive.
- If result<0, review is negative.
- If result=0, review is neutral.

### VIII. PROPOSED ALGORITHM

1. Read the input sentence.
2. Parse the input into tokens and store in string array.
3. Identify the word as sentiment word or noun word in the paragraph using database.
4. Set: -position\_noun=-1, position\_property\_noun=-1, position\_noun\_diff=0, position\_sentiment=0, diff=0, temp=0, result=0, result1=0, result2=0, ans=0.
5. If ( word = noun)
  - If( position\_noun < 0)
    - position\_noun = position\_word
  - else
    - position\_property\_noun=position\_noun
6. If (word = sentiment)
  - position\_seniment = position\_word

7. position\_noun\_diff=position\_noun-position\_property\_noun
  - if (position\_noun\_diff < 0)
    - diff=position\_property\_noun- position\_noun
  - temp = position\_property\_noun;
    - else if(position\_noun\_diff > 0)
      - diff=position\_nounposition\_sentiment;
      - temp = position\_noun;
8. Extract the sentiment score of sentiment word from corpus and save in result.
9. Find the position of negation word:
  - a) If the difference of position of sentiment word and position of negation word is 1, then: result= -result
  - b) Else: result = result
10. If more than one sentiment word is found, answer is updated as:
  - ans = ans + result
11. result1 = ans
12. Case A for AND conjunction
  - a) Conjunctive word और/एवं found
    - If (position of sentiment word < position of conjunctive word and position of sentiment word > position of negation word)
      - ans = -ans
  - b) Sentiment word is present before conjunction and after conjunction
    - i. Extract the score of first part of sentence (before conjunction) and save in result1.
    - ii. Extract the score of second part of sentence (after conjunction) and save in result2.
    - iii. If (result1 > 0 and result2 < 0)
      - result2 = -result2
      - else if (result1 < 0 and result2 > 0)
        - result2= -result2
- Case B for BUT conjunction
  - a) Conjunctive word लेकिन/पर/मगर/परंतु found
    - If (position of sentiment word < position of conjunctive word and position of sentiment word > position of negation word)
      - ans = -ans
  - b) Sentiment word is present before conjunction and after conjunction
    - i. Extract the score of first part of sentence (before conjunction) and save in result1.
    - ii. Extract the score of second part of sentence (after conjunction) and save in result2.
    - iii. If (result1 > 0 and result2 > 0)
      - result2 = -result2
      - else if (result1 < 0 and result2 < 0)
        - result2= -result2
13. ans = result1 + result2
14. If ( ans > 0)
  - if (-3 <= diff && diff <= +3 && diff != 0)
    - if (position\_noun\_diff < 0)
      - then: Hindi word present at position\_property\_noun is an aspect of the word present at position\_noun. Also, the word present at position\_sentiment is used to indicate the

entity present at position\_property\_noun. Also, the review is positive.

Else

Display: The Hindi word present at position\_sentiment is used to indicate the entity present at position\_property\_noun. The review is positive.

Else

Display: review is positive

15. If ( ans < 0)

if (-3 <= diff && diff <= +3 && diff != 0)

if (position\_noun\_diff < 0)

then: Hindi word present at position\_property\_noun is an aspect of the word present at position\_noun. Also, the word present at position\_sentiment is used to indicate the entity present at position\_property\_noun. Also, the review is negative.

Else

Display: The Hindi word present at position\_sentiment is used to indicate the entity present at position\_property\_noun. The review is negative.

Else

Display: review is negative

16. If ( ans = 0)

Display: review is neutral.

17. EXIT

## IX. RESULTS

Purpose of the analysis is to convert unstructured data into some meaningful information. Once the analysis is completed, numbers of alternatives are used to display the result of text analysis [23]. Opinions on various aspects (scooter, design, and seat) of ActivaHonda scooter collected in database are represented in fig 3.

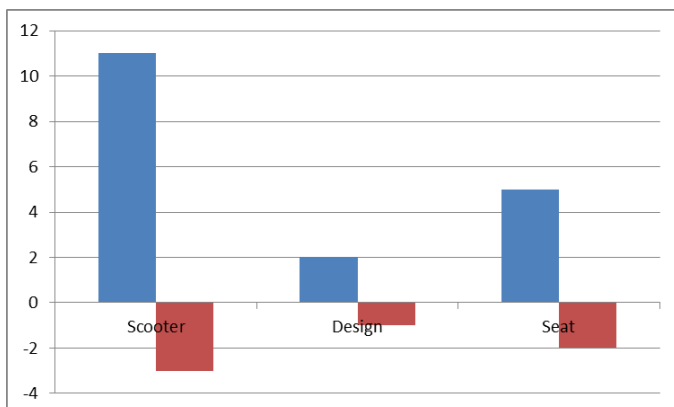


Figure 3

System is tested for performance analysis [24]. The results are computed after crawling 210 reviews from different websites. The results for the experiments are conducted on 110 positive and 100 negative reviews. When algorithm is applied to reviews using simple case, out of 210 reviews- 60 reviews were correctly positive and 48 reviews were correctly negative. Remaining 102 reviews were showing false output. When algorithm is applied to reviews using negation handling, out of 210 reviews- 85 reviews were correctly positive and 72

reviews were correctly negative; remaining 53 reviews were showing false output. Similarly, when algorithm is applied to reviews using negation handling and conjunction handling, out of 210 reviews- 96 reviews were correctly. With improvement in algorithm, the system is repeated to verify its results and there is significant increase in accuracy [24]. Following graph shows the accuracy of system (in different cases):

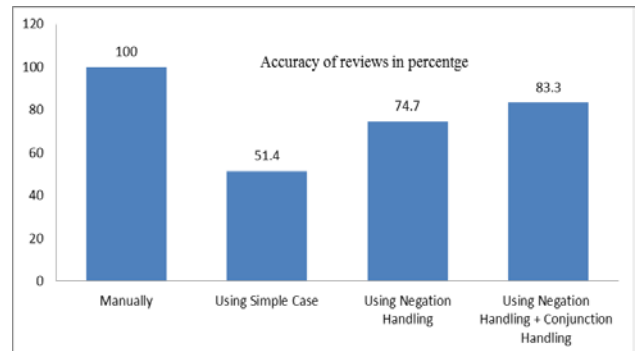


Figure 4

## X. CONCLUSION AND FUTURE SCOPE

Sentiment Analysis is an emerging field and is important as human beings are largely dependent on it nowadays. Its use has been spread to services and making applications and advancements in this area. SA has led to building of better products, understanding user's opinions, executing and managing of business decisions. Target of SA is to make computer able to identify and create emotions like a human being. One of the biggest challenges in SA for Hindi language is the scarcity of resources. This work aims to create a data bank to facilitate the referencing needs of researchers and practitioners in this area. We designed an algorithm which extracts the sentiment (or opinion) from a review sentence and also the object for which the opinion is for. It also calculates the sentiment score of the sentence using database dictionary.

This thesis focuses on identification of sentiment words, noun, negation words, conjunction words and sentiment score calculation using negation handling and conjunction handling.

In the future, application developers should work on the other aspects like adjectives and verbs with noun. In our research we identify position conjunctive, negative and noun words to calculate the results. In the future, researcher solved the problem of sentiment analysis including POS tagging.

## XI. REFERENCES

- [1] Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.
- [2] Shailendra Kumar Singh and Sanchita Paul, "Sentiment Analysis of Social Issues and Sentiment Score Calculation of Negative Prefixes", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.55 (2015)
- [3] Yao, R., & Chen, J. (2013), "Predicting Movie Sales Revenue using Online Reviews", International Conference on Granular Computing (GrC), IEEE.
- [4] Yung-Ming, L., & Tsung-Ying, L. (2013), "Deriving market intelligence from micro-blogs Decision Support System", Decision Support Systems 55 (2013) 206-217.

- [5] Bo Pang and Lillian Lee, "Opinion Mining and Sentiment Analysis".
- [6] Singh, V., & Piryani, R. (2013), "Sentiment Analysis of Movie Reviews: A new feature-based heuristic for aspect-level sentiment classification", Department of Computer Science, South Asian University, New Delhi.
- [7] Joshi, A., A R, B., & Bhattacharyya, P. (2010), "A fall strategy for sentiment analysis in Hindi : A case study", International Conference on Natural Language Processing.
- [8] A. Joshi, B.A.R, and P.Bhattacharyya, "A fall-back strategy for sentiment analysis in Hindi: a case study", In International Conference On Natural Language Processing (ICON), 2010.
- [9] Balamurali, A. R., Joshi, A., and Bhattacharyya, P. (2012), "Cross-lingual sentiment analysis for indian languages using linked wordnets", In COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Posters, 8-15 December 2012, Mumbai, India, pages 73-82.
- [10] Balamurali, A. R., Joshi, A., and Bhattacharyya, P. (2011), "Harnessing wordnet senses for supervised sentiment classification", in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, 27-31 July 2011, John McIntyre Conference Centre, Edinburgh, UK, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1081-1091.
- [11] Bakliwal, A., Arora, P., and Varma, V. (2012), "Hindi subjective lexicon: A lexical resource for hindi polarity classification", Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC).
- [12] Mittal, N., Agarwal, B., Chouhan, G., Bania, N., and Pareek, "Sentiment analysis of hindi review based on negation and discourse relation", In proceedings of International Joint Conference on Natural Language Processing, 2013, pages 45-50.
- [13] Sharma, R., Nigam, S., and Jain, R. (2014), "Polarity detection movie reviews in hindi language", CoRR, abs/1409.3942.
- [14] Das, A. and Bandyopadhyay, S. (2010a), "Phrase-level polarity identification for bangla", Int. J. Comput. Linguist, Appl.(IJCLA), 1(1-2):169-182.
- [15] Walla Medhat, Ahmed Hassan and Hoda Korashy, "Sentiment Analysis Algorithms and Applications: A Survey", Ain Shams Engineering Journal (2014) 5, 1093-1113.
- [16] Miller G, Beckwith R, Fellbaum C, Gross D, Miller K," WordNet: an on-line lexical database", Oxford Univ. Press; 1990.
- [17] KIM, S. and HOVY, "Determining the sentiment of opinions", In Proceedings of the 20th international conference on Computational Linguistics, 1367-1373, 2004.
- [18] Guang Qui, Xiaofei He, Feng Zhang and C.L. Philip Chen, "DASA: Dissatisfaction-oriented Advertising based on Sentiment Analysis", Expert Systems with Applications 2010; vol. 37:6182-91.
- [19] B. Liu. Sentiment Analysis and Subjectivity, "Handbook of Natural Language Processing", Second Edition, (editors: N. Indurkha and F. J. Damerau), 2010. ....
- [20] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), pp. 168-177, 2004.
- [21] G.Vinodhini and RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012.
- [22] Tanveer Khan and Yogesh Kumar, "A Novel Approach to Analyze the Sentiment with Conjunctive Words", International Research Journal of Engineering and Technology, Volume: 03 Issue: 07, July-2016
- [23] Meena Rambocas and Joao Gama, "Marketing Research: The Role of Sentiment Analysis", FEP Working Papers.
- [24] Amandeep Kaur and Vishal Gupta, "A Survey on Sentiment Analysis and Opinion Mining Techniques", Journal Of Emerging Technologies In Web Intelligence, Vol. 5, No. 4, November 2013.