International Journal of Advances in Computer Science and Technology

Determination of Polarity of Sentences using Sentiment Orientation System

Richa Sharma[1], Shweta Nigam[2], Rekha Jain[3]
[1]M.Tech Scholar, India, sharma.richa3889@gmail.com
[2]M.Tech Scholar, India, nigam.1611@gmail.com
[3]Assistant Professor, India, Rekha_leo2003@rediffmail.com

ABSTRACT

Opinions are always very important for human beings. Whenever a decision has been carried out, people always consult with the friends and relatives. But from the last few years , the impact of the web has surprisingly increased, for peoples web documents act as a new source of opinion. Nowadays each company allows its customers to write their opinions related to their products, so large numbers of customer opinions i.e. reviews are available on the Web related to every product. To analyze this large amount of information it is required to develop a method that automatically classifies these reviews. Opinion Mining or Sentiment Analysis is the mining of opinions and emotions automatically from text through Natural Language Processing (NLP). In this paper an opinion mining system named as "Sentiment orientation System" is proposed using unsupervised technique to determine the polarity of sentences. Negation is also handled in the proposed system. Experimental results using reviews of products show the effectiveness of the system.

Key words : Customer Reviews, Opinion Mining, Sentiment Analysis,Wordnet.

1. INTRODUCTION

Opinion Mining is an emerging area of research nowadays, as the influence of the Web is increasing among people, internet is reaching to everyone's in this world. Now peoples would like to express their opinion, feelings and experiences on the web. There are lots of sources available on the web through which people share their opinions and experiences. People commonly use blogs and social networking sites like twitter, facebook etc. to express their views and opinions. Companies also allow their customers to express their opinion on their websites in the customer review column, these opinions help the organizations in improving their services as they know about what peoples think about their products. People expresses opinion each and every day so large amount of user data is available on the Web, thus mining the data and identifying user sentiments, likes and dislikes is one of an important task. Opinion Mining or Sentiment Analysis is a natural language processing task that mine information from various text forms such as reviews, news, and blogs and classify them on the basis of their polarity as positive, negative or neutral. Opinion Mining categorizing the text as subjective and objective. Subjectivity means that the text

contains opinion whereas Objectivity means that the text is without opinion .
For examples-
1. **Subjective-** *This movie by Shahrukh Khan and Kajol is superb.*(this sentence contains opinion, it talks about the movie and the writer's feelings about word "superb" and hence it's subjective).

The subjective text can be further divided into 3 categories based on the sentiments expressed in the text.
1. **Positive**- *I love to watch Indian movies.*
2. **Negative**- *The movie was disaster.*
3. **Neutral**- *This is my book.* (this sentence has user's views, feelings hence it is subjective but as it does not have any positive or negative polarity so it is neutral .)

2. **Objective-** *This movie stars Shahrukh Khan and Kajol.*(this sentence is a fact, does not contain any opinion, hence its objective)

1.1 Components of Opinion Mining

There are mainly three components of Opinion Mining, these are:
1. Opinion Holder: Opinion holder is the person or organization that expresses the opinion.
2. Opinion Object: It is a feature about which the opinion holder is expressing his opinion.
3. Opinion Orientation: Determine the orientation of the opinions expressed in the text.
For example "This movie has an excellent story". In this review, the person who has given the opinion is the opinion holder. Opinion object here is the *story* of the movie and the opinion word is "excellent" which is positively orientated. Semantic orientation is a task of determining whether a sentence has either positive, negative orientation or neutral orientation [9] [10].

1.2 Levels of Opinion Mining

Opinion mining is performed at one of these levels[7]-
1) Document Level- The document level classifies the whole document as a single polarity positive, negative or neutral.
2) Sentence Level -The sentence level analyze the documents at sentence level. The sentences are analyzed individually and the corresponding polarity is determined.
3) Aspect Level- The aspect level analysis going much deeper and deals with identifying the features in a sentence for a

given document and analyze the features and classify them as positive, negative or neutral.

In this paper, an opinion mining system named as "Sentiment orientation system" is proposed which extracts the opinions from the given text and determines whether the given text is positive, negative or neutral. Negation is also handled in this system. To determine the semantic orientation of the sentences a dictionary based technique of the unsupervised approach is adopted. Rest of the paper is organized as follows: Section 2 describes the related work .Section 3 describes the proposed approach. Section 4 compares the experimental results of this system with the other system and it shows that the proposed system gives better result. Section 5 concludes the paper.

2. EXISTING RESEARCH WORK

In the field of Opinion Mining, The most promising work had done by Turney [12].He used unsupervised method for polarity detection . Two seed words "poor" and "excellent" were taken to determine the semantic orientation of the phrases, point wise mutual information method was used to determine the semantic orientation.66% accuracy was achieved in his experiment for the movie review domain.

Ting-Chun Peng [14] extracts the sentiment from each review by using part-of-speech (POS) patterns, as a query term unknown sentiment phrase was used and top-N relevant phrases from a search engine were extracted. After that, sentiments of these unknown phrases are computed based on the sentiments of nearby known relevant phrase using lexicons.

Andrea Esuli [1] proposed a semi-supervised learning method, in this method they started from expanding an initial seed set using WordNet. To determine the semantic orientation of the expanded seed terms statistical technique was used.

Chunxu Wu [5] proposed an approach in which to judge the orientation of opinion, semantic similarity measures were used. Orientation of context independent opinions was determined and by using linguistic rules, orientation of context distinct-dependent opinion were considered .After that, contextual information from other reviews that comment on the same product feature to judge the context indistinct-dependent opinions were extracted.

Zagibalov and Carroll [15] proposed a method of automatic seed word selection for unsupervised sentiment classification of product reviews in Chinese.

3. PROPOSED SYSTEM

Unsupervised technique is used in this system. This technique does not require prior training of data in order to mine the data [2].Among the various available methods in the unsupervised

technique, dictionary based approach is used in this system to determine the orientation of sentences. WordNet [6] is used as a dictionary to determine the opinion words and their respective synonyms and antonyms. The proposed work is closely related to the Minqing Hu and Bing Liu work on Mining and Summarizing Customer Reviews [8]. Fig1. gives the overview of the proposed system 'Sentiment Orientation System'[13]. Customer reviews of the mobile phones are collected from the Amazon website (www.amazon.in) which is used as an input to the system. Output is that the system generate the separate cluster and graphical output of positive, negative and neutral sentences which will be helpful to the user to read ,analyze and help them in taking the decision whether the product is to be purchased or not. The polarity of the given sentence is determined on the basis of the majority of opinion words.

The work is divided into following steps:-

1) **Preprocessing**

   To determine the polarity large number of reviews are collected, these reviews are first preprocessed and then sent to the POS tagger [4][7]which gives the tagged output, POS tagging is necessary to identify the opinion words.

2) **Seed List Preparation**

   Initially some of the common opinion words along with their polarity are stored in the seed list .After that opinion words are extracted from the given reviews and their synonyms and antonyms are determined with the help of WordNet and matched with the words stored in seed list, if it matched with the word in positive orientation then the word is store with the same polarity otherwise it is stored with the opposite polarity. In this way the seed list keep on increasing. It grows every time whenever the new antonyms and synonyms words are found in WordNet matches with seed list.

3) **Polarity Determination**

   With the help of seed list, the polarity of the sentences is determined. Polarity is determined on the basis of majority of opinion words, if the number of positive words are more ,then the sentence is positive otherwise the sentence is negative and if the number of positive and negative words are equal then the sentence shows the neutral polarity. Negation is also handled in the system, if the opinion word is preceded by negation "not" then the polarity of that sentence is reversed.
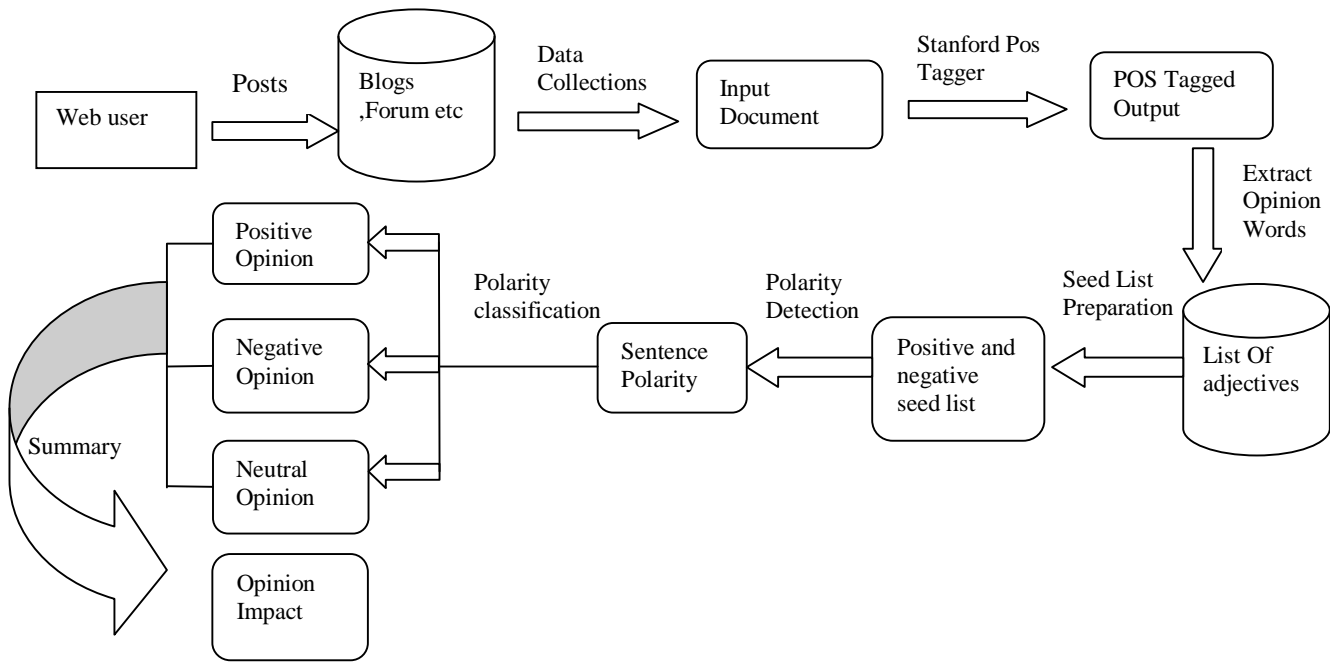
Figure 1. Sentiment Orientation System

4. EXPERIMENTS & RESULTS

The experiment is conducted using the customer reviews of phones. All the reviews were collected from the Amazon.com (www.amazon.in).This website contains a large number of customer reviews of various products. Initially, reviews of the mobile phones has been collected and applied to the proposed system. The result shows the orientation of each collected review . The final results are shown in graphical charts. As we all know human always gives the correct opinion so to evaluate how accurate the proposed system is as compared to the human decision, we manually read all the sentences and determine whether the sentences are positive, negative or neutral and then compare these results with the results of 'Sentiment orientation System'. Apart from phone domain the experiment is also performed on the other domains to determine the system's performance in different domains. Domains that are considered are:

- Watch review Domain
- Bollywood reviews Domain
- Books reviews Domain

Three evaluation measures are used on the basis of which systems are compared, these are:-

- Precision
- Recall
- Accuracy

Accuracy is the portion of all true predicted instances against all predicted instances. An accuracy of 100% means that the predicted instances are exactly the same as the actual instances.

$$\text{Accuracy} = \frac{(tp+tn)}{tp+tn+fp+fn} \qquad (1)$$

Precision is the portion of true positive predicted instances against all positive predicted instances.

$$\text{Precision} = \frac{tp}{tp+fp} \qquad (2)$$

Recall is the portion of true positive predicted instances against all actual positive instances.

$$\text{Recall} = \frac{tp}{tp+fn} \qquad (3)$$

The common way for computing these indexes is based on the confusion matrix shown in Table 1.

Table 1. Confusion Matrix

Instances	Predicted positives	Predicted negatives
Actual positive instances	# of True positive instances (TP)	# of false negative instances (FN)
Actual negative instances	# of false positive instances (FP)	# of True Negative instances (TN)

On the basis of these evaluation measures, results show that 'Sentiment orientation System' is performed well in all the domains that are considered. The experiments have been performed on each domain (with 50 sentences) collected from the Amazon website.
Table 2 and fig.2 presents the results of 'Sentiment Orientation System' on Phone reviews domain.
Table 3 and fig.3 presents the results of 'Sentiment Orientation System 'on Watch reviews domain'.
Table 4 and fig.4 presents the results of 'Sentiment Orientation System 'on Bollywood reviews domain'.
Table 5 and fig.5 presents the results of 'Sentiment Orientation System 'on Books review domain'.

Table 6 and fig.6 represents the comparison of the results of the all the domain which shows that performance of 'Sentiment Orientation System' is good in all the domains considered
.
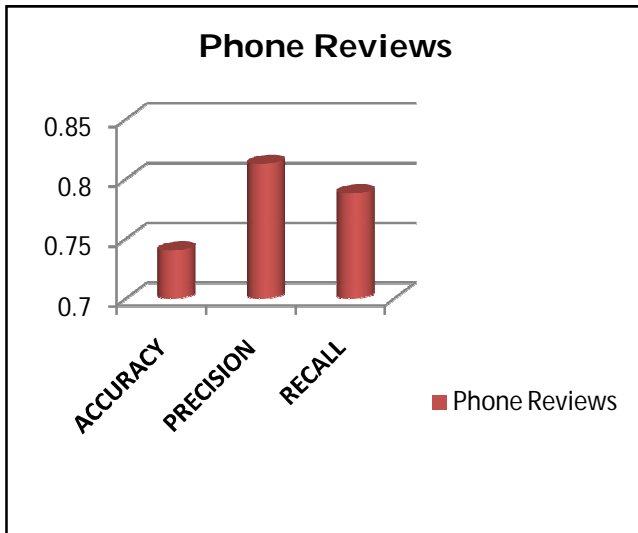'Sentiment Orientation System' performs well with respect to all the domains which prove that system is efficient.

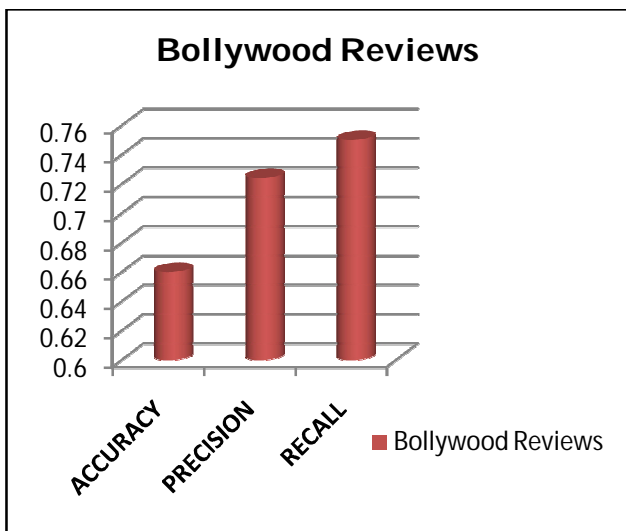**Figure 2**. 'Sentiment Orientation System' graph on phone domain

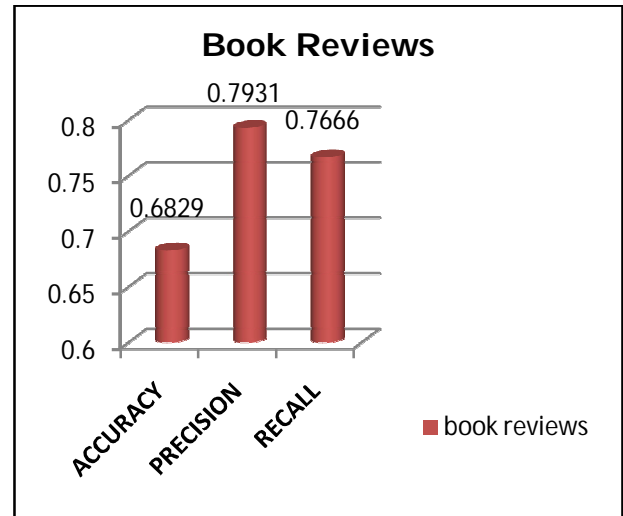**Figure 4**. 'Sentiment Orientation System' graph on book domain

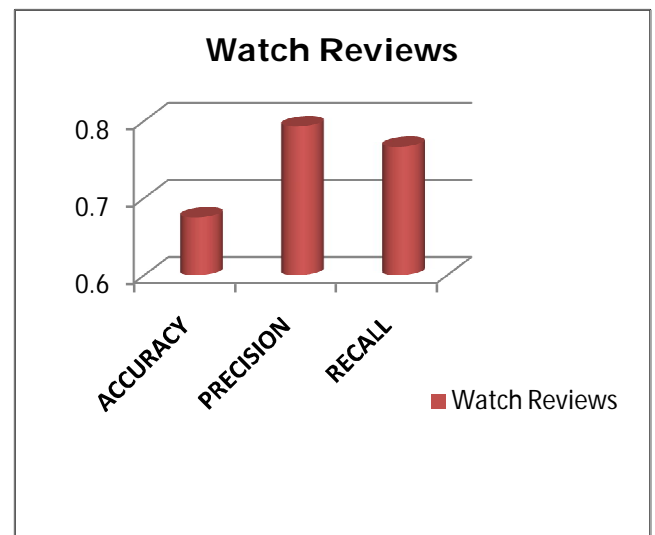

 Figure 5. 'Sentiment Orientation System' graph on watch domain

Figure 3. 'Sentiment Orientation System' graph on bollywood domain

Table 2 Final comparison of all the domains

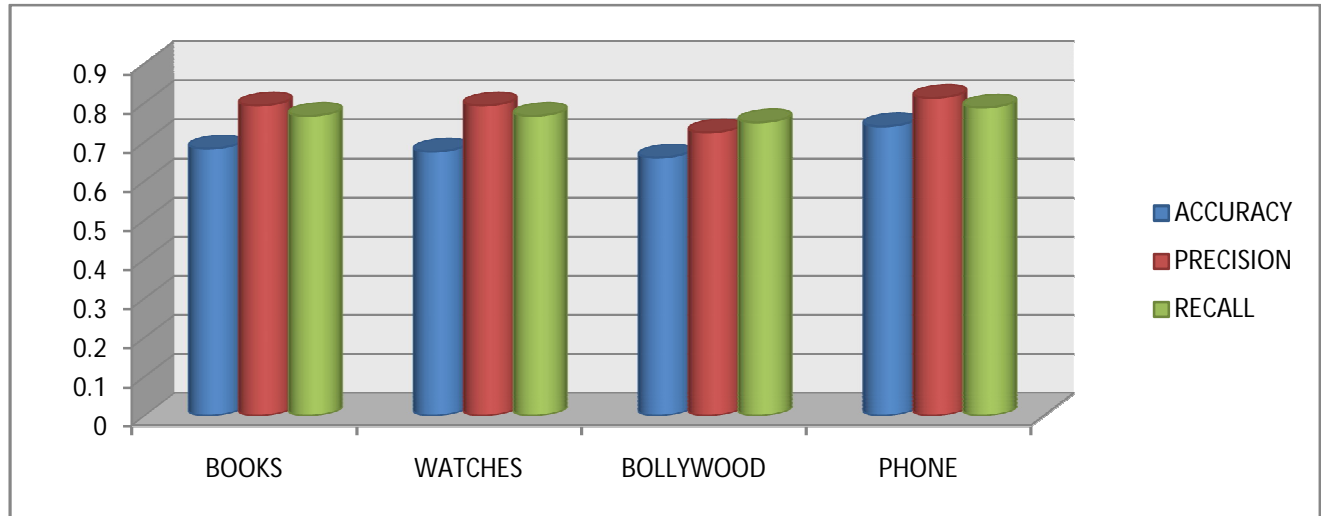| Table Measures\Domain	Books	Watches	Bollywood	Phone
Accuracy	0.6829	0.6751	0.6601	0.7401
Precision	0.7931	0.7931	0.7241	0.8125
Recall	0.7666	0.7666	0.7501	0.7878

Figure 6. Graphical representation of all the domains

5. CONCLUSION

The objective of this paper is to determine the semantic orientation of a large number of customer reviews for the product. The summarized results generated by the system will be helpful for the user in taking the decision .Experimental results indicate that the 'Sentiment Orientation System' performs well in all the domains that are considered. This task is really important because nowadays people are largely dependent on the opinions expressed on the web. This is also helpful for the enterprises as they easily determine the pros and cons of their product. In future, efforts would be done to make some enhancements in this technique. It would deal with the sentences contain relative clauses like not only-but also and the sentences contain clauses neither-nor, either-or etc. it would determine the polarity based on the aspects of the sentences.

REFERENCES

1. Andrea Esuli and Fabrizio Sebastiani. **Determining the semantic orientation of terms through gloss classification**.
2. B. Pang and L. Lee. **Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval**, 2(1-2),1-135,2008.
3. Bing Liu. **Sentiment Analysis and Opinion Mining**, *Morgan & Claypool Publishers,*May 2012.
4. Christopher D. Manning. **Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?** *Published in CICLing'11 Proceedings of the 12th international conference on Computational linguistics and intelligent text processing,*Volume Part I.
5. Chunxu Wu, Lingfeng Shen. **A New Method of Using Contextual Information to Infer the Semantic Orientations of Context Dependent Opinions**.
6. George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. **Introduction to Word Net: An On-line Lexical Database (Revised August 1993)** International *Journal of Lexicography 3(4):235-244* (*1990*).
7. Kevin Gimpel, Nathan Schneider, Brendan O 'Connor, Dipanjan Das, Daniel Mills,Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. **Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments**, *Published in Proceeding HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers -,* Volume 2 , ISBN: 978-1-932432-88-6.
8. M. Hu and B. Liu.**Mining and summarizing customer reviews**, *presented at the Proceedings of the tenth ACM.*
9. Mohammad Sadegh Hajmohammadi, Roliana Ibrahim, Zulaiha Ali Othman .**Opinion Mining and Sentiment Analysis: A Survey**, ISSN: 2277–3061 (online) . *International Journal of Computers & Technology* ,Volume 2 ,No. 3, June, 2012.
10. Nilesh M. Shelke, Shriniwas Deshpande, Vilas Takre. ,**Survey of techniques for opinion mining**, *International Journal of Computer Applications,* 57**,**13,pp 0975-8887,2012.
11. Pang, B., Lee, L., & Vaithyanathan, S. **Thumbs up? Sentiment classification using machine learning techniques,** *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pp 79-86.
12. P. Turney, **Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of**

reviews, *Proceedings of the Association for Computational Linguistics*.

13. Richa Sharma, Shweta Nigam, Rekha Jain. **Polarity detection at sentence level**. *International journal of computer applications(0975 – 8887),* Volume 86 – No 11, January 2014

14. Ting-Chun Peng, Chia-Chun Shih. **Using Chinese part-of-speech patterns for sentiment phrase identification and opinion extraction in user generated reviews, Digital Information Management (ICDIM).**

15. Zagibalov, Taras, and John Carroll. **Unsupervised Classification of Sentiment and Objectivity in Chinese Text.** *In Proceedings of the Third International Joint Conference on Natural Language Processing.* 304–311,2008.