

Polarity Detection at Sentence Level

Richa Sharma
M.Tech Scholar
Banasthali Vidyapith C-62
Sarojini marg C-Scheme India

Shweta Nigam
M.Tech Scholar
Banasthali Vidyapith C-62
Sarojini marg C-Scheme India

Rekha Jain
Assistant Professor
Banasthali Vidyapith C-62
Sarojini marg C-Scheme India

ABSTRACT

Opinions bear a very important place in the life of human beings. Human Beings are always surrounded by opinions, when a decision has to be taken; people always want to know the opinions of others. But as the impact of the web is increasing day by day, Web documents can be seen as a new source of opinions for the people. Large numbers of reviews are available on the Web related to every product. Whenever a customer buys any product, they express their feedbacks as opinions on the e-commerce website, thus it is very important to automatically analyze the huge amount of information on the web and develop methods to automatically classify the reviews. Opinion Mining or Sentiment Analysis is the mining of attitudes, opinions, and emotions automatically from text, speech, and database sources through Natural Language Processing (NLP). In this paper an opinion mining system is proposed using unsupervised technique to determine the polarity of sentences i.e. to classify the sentences as positive, negative or neutral. Negation is also handled in the proposed system. Experimental results using reviews of products show the effectiveness of the system.

Keywords

Opinion Mining, Sentiment Analysis, Reviews, WordNet.

1. INTRODUCTION

Opinions and emotions play an important role in human life. Opinions influenced the thinking of human beings as humans always want to know the opinions of others in every decision which they made. Opinion can be defined as “A view or judgment formed about something either made by an individual or a majority of people”. Nowadays, people spend their most of the time on the web, surfing on the Internet is common now, and web documents act as a new source of experiences and individual opinions. Opinions are available for every product on the web as people’s buy and use the product, they express their opinions and experiences on the respective websites and these opinions help the people to determine whether that product is suitable for them or not. Millions of opinions are available on the web so it is important to automatically analyze the large amount of information and develop the method that automatically classifies these opinions on the web. This new research domain is usually called Opinion Mining and Sentiment Analysis and it increases the huge interest in researchers. Opinion Mining can be defined as a sub-discipline of computational linguistics that focuses on extracting opinion of persons from the web. It is a *Natural Language Processing (NLP)* and *Information Extraction (IE)* task that aims to obtain feelings of the writer expressed in positive or negative comments by analyzing a large number of documents [10]. It combines the techniques of computational linguistics and Information Retrieval (IR).

Three main components of Opinion Mining 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 whether the opinion about an object is positive, negative or neutral.

For example “This television has an excellent *picture quality*”. In this review, Opinion Holder is the user who has written this review. Opinion object here is the *picture quality* of the television 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 [11, 12].

Sentiment analysis tasks mainly consist of classifying the polarity of a given text at the document, sentence or feature/aspect level. It expresses polarity as positive, negative or neutral. The sentiment analysis can be performed at one of the three levels: the document level, sentence level, feature levels [3]

- Document level: This level classifies the whole opinion document as positive, negative or neutral. The task performed at this level is commonly known as document-level sentiment classification.
- Sentence level: This level classifies the sentences and determines whether the sentences indicate positive, negative or neutral opinion. The task performed at this level is commonly known as sentence-level sentiment classification.
- Entity and Aspect level: This level classifies sentences/documents based on the aspects of those sentences. It identifies the aspects from sentences and determines the orientation of each aspect. The task performed at this level is commonly known as aspect-level sentiment classification.

An opinion mining system named as “Sentiment orientation system” is proposed which extracts the opinions from sentences and determines whether the given sentences are positive, negative or neutral. The system also handles the negative sentences and gives the appropriate polarity of the sentences. A dictionary based technique of the unsupervised approach is adopted to determine the semantic orientation of the sentences. WordNet is used as a dictionary to determine the respected opinion words and their synonyms and antonyms. The rest of the paper is organized as follows: Section 2 discusses 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

Much of the research in unsupervised sentiment classification makes use of lexical resources which are available on the web. Turney [12] had done the most prominent work by using unsupervised methods for Opinion Mining and sentiment detection. Two seed words “poor” and “excellent” were taken

because they are mostly present on the web to calculate the semantic orientation of the phrases, orientation is measured by point wise mutual information. The sentiment of a document is calculated as the average semantic orientation of all such phrases. 66% accuracy was achieved in his experiment for the movie review domain.

Ting-Chun Peng [13] extracts the sentiment phrases of each review by using **part-of-speech (POS) patterns**, unknown sentiment phrase as a query term was used and top-N relevant phrases from a search engine were extracted. After that, sentiments of unknown sentiment 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. Statistical technique was used to determine the semantic orientation of the expanded seed terms.

Chunxu Wu [5] proposed an approach in which semantic similarity measures were used to judge the orientation of opinion. Orientation of context independent opinions was determined and the context dependent opinions using linguistic rules to infer 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 [14] proposed a method of automatic seed word selection for unsupervised sentiment classification of product reviews in Chinese.

3. PROPOSED SYSTEM

This system is based on unsupervised technique. 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 to determine the orientation of sentences. WordNet [6] is used as a dictionary to determine the opinion words and their synonyms and antonyms. The proposed work is closely related to the Mingqing Hu and Bing Liu work on Mining and Summarizing Customer Reviews [8]. Fig1. gives the overview of the proposed system 'Sentiment Orientation System'. Customer reviews of the mobile phones are collected from the Amazon website (www.amazon.in) which is used as an input to the system. The system will generate the cluster of positive, negative and neutral sentences which will be easier for users 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.

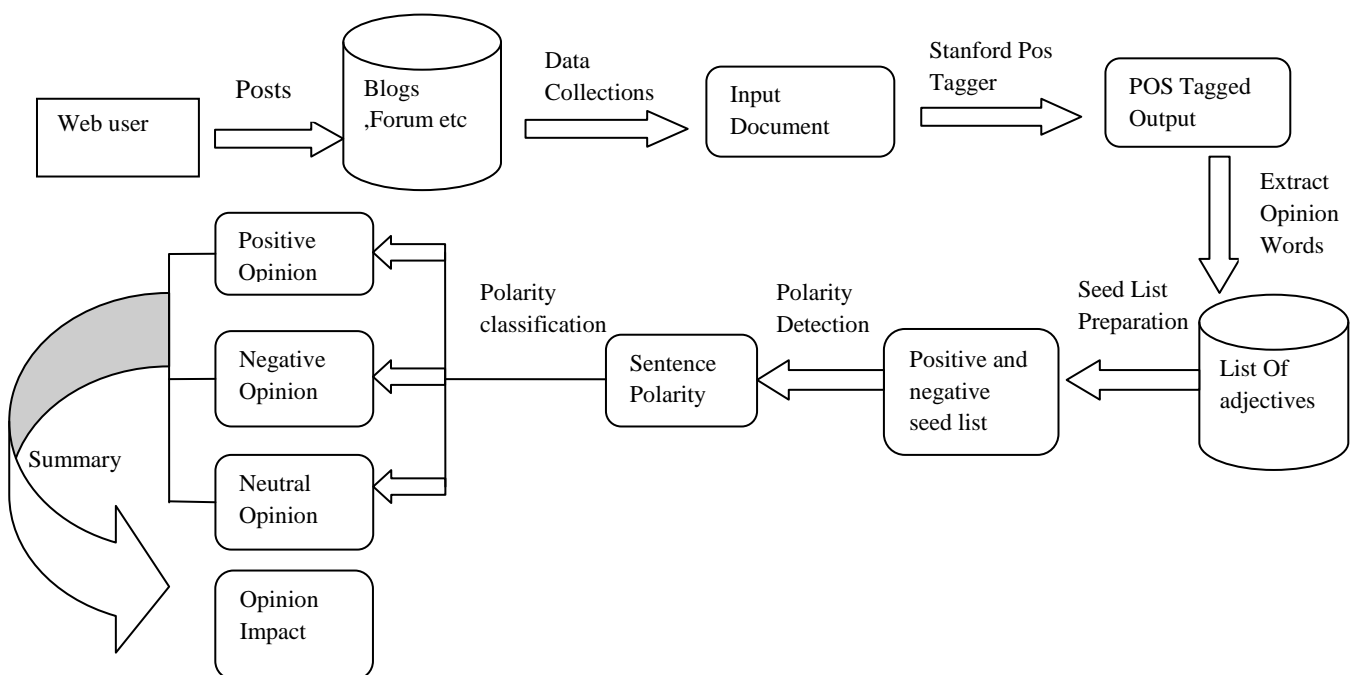


Fig 1: Sentiment Orientation System

The system performs this task in several steps as follows:-

1) Preprocessing

All the reviews that have been collected are preprocessed first 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

It consists of two clusters of the positive and negative words. Some of the common opinion words are stored in the seed list and for all the extracted opinion words, synonyms and antonyms are determined with the help of WordNet and matched with the words stored in the seed list if it matched with the word in positive cluster then the word is added into the positive cluster otherwise it is added in the negative cluster. In this way the seed list keeps 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 are more, then the sentence is positive otherwise 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.

4. EXPERIMENTS & RESULTS

The experiment is conducted using the customer reviews of mobile 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 sentence i.e. whether a sentence is positive, negative or neutral. 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’. The same work has been conducted on another system ‘AIRC Sentiment Analyzer’, which is available online. Finally the results are compared and it is found that the performance of the proposed system is better than the ‘AIRC Sentiment Analyzer’. Three evaluation measures are used on the basis of which systems are compared, these are:-

- Precision
- Recall
- Accuracy

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)

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.

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

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

$$\text{precision} = \frac{tp}{tp+fp} \quad (2)$$

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

$$\text{recall} = \frac{tp}{tp+fn} \quad (3)$$

On the basis of these evaluation measures, results show that ‘Sentiment orientation System’ is more efficient as compared to the ‘AIRC Sentiment Analyzer’. The experiments have been performed on both systems using 50 sentences of phone reviews which were collected from the Amazon website.

Table 2 and fig.2 presents the results of ‘Sentiment Orientation System’.

Table 3 and fig.3 presents the results of ‘AIRC Sentiment Analyzer’.

Table 4 and fig.4 represents the comparison of the results of the two systems which shows that performance of ‘Sentiment Orientation System’ is better than the ‘AIRC Sentiment Analyzer’.

‘Sentiment Orientation System’ performs well as compared to the ‘AIRC Sentiment Analyzer’ System with respect to all the evaluation measures. ‘Sentiment orientation system’ shows the accuracy of 74% which proves the system more efficient.

Table 2. ‘Sentiment Orientation System’ Results

Measures	Results
Accuracy	0.7401
Precision	0.8125
Recall	0.7878

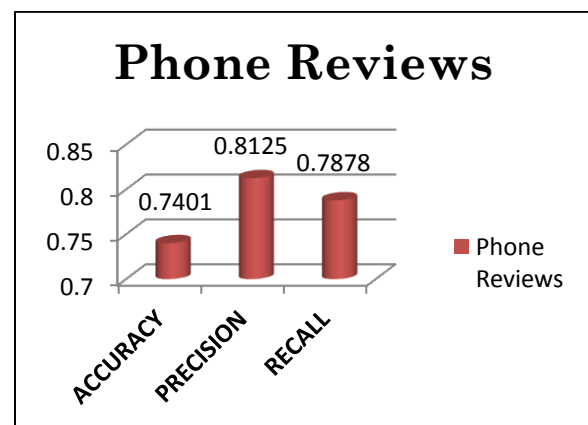


Fig2. ‘AIRC Sentiment Analyzer’ Graph

Table 3. ‘AIRC Sentiment Analyzer’ Results

Measures	Results
Accuracy	0.7011
Precision	0.8125
Recall	0.6841

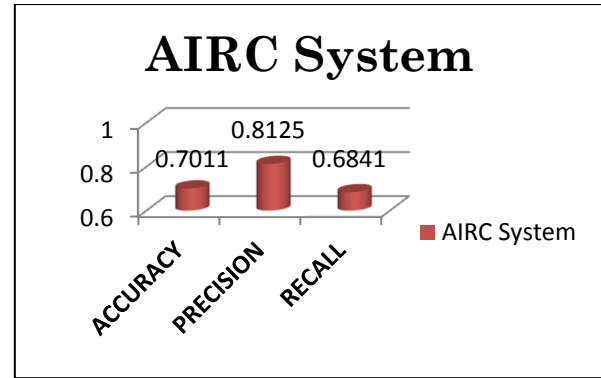


Fig3. ‘AIRC Sentiment Analyzer’ Graph

Table 4. Comparison of Sentiment Orientation System & AIRCS

Measures\System	AIRC	Sentiment Orientation System
Accuracy	0.7011	0.7401
Precision	0.8125	0.8125
Recall	0.6841	0.7878

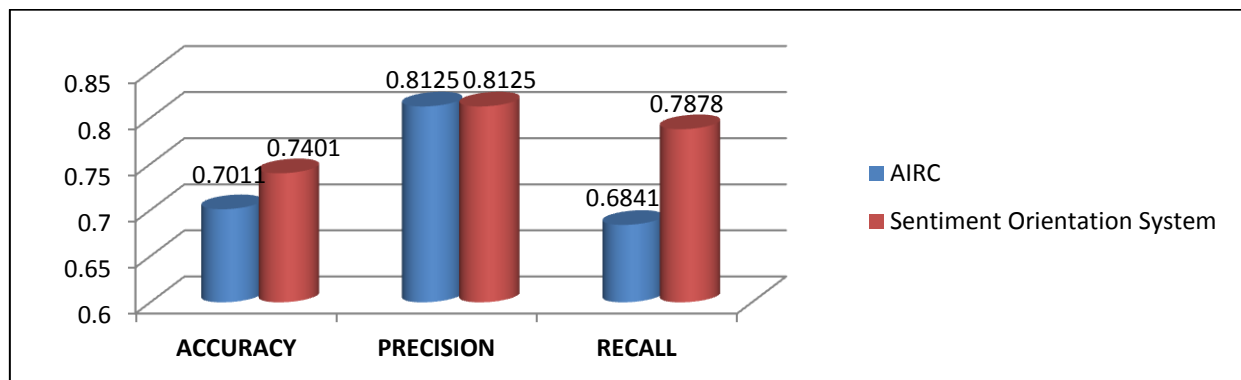


Fig4. Comparison of Sentiment Orientation System & AIRCS Sentiment Analyzer

5. CONCLUSION

The objective of this paper is to determine the polarity 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 better than the ‘AIRC Sentiment Analyzer System’. This task is really important because nowadays people are mostly dependent on the opinions expressed on the web. This is also helpful in the organization as they easily determine the pros and cons of their product. In future work, 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. We also plan to determine the polarity based on the aspects of the sentences.

6. 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 WordNet: 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(2012), Opinion Mining and Sentiment Analysis: A Survey, ISSN: 2277–3061 (online). International Journal of Computers & Technology Volume 2 No. 3, June, 2012.
- [10] NileshM.Shelke, ShriniwasDeshpande, Vilas Takre (2012), Survey of techniques for opinion mining,International Journal of Computer Applications, 57,13,pp 0975-8887.
- [11] Pang, B., Lee, L., & Vaithyanathan, S. (2002). 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] 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).
- [14] Zagibalov, Taras, and John Carroll. 2008. Unsupervised Classification of Sentiment and Objectivity in Chinese Text. In Proceedings of the Third International Joint Conference on Natural Language Processing. 304–311.