

# Sentiment Analysis

Prof. Reena Mahe<sup>1</sup> Prof. Sumita Chandak<sup>2</sup> Prof. Nileema Pathak<sup>3</sup> Prof. Renuka Nagpure<sup>4</sup> Prof. K. Tejaswi<sup>5</sup>

<sup>1</sup>Lecturer <sup>2,3,4,5</sup>Assistant Professor  
<sup>1,2,3,4,5</sup>Department of Information Technology  
<sup>1,2,3,4,5</sup>Atharva College of Engineering University of Mumbai, India

*Abstract*— With the evolution of web technology and social media, online shopping has become very famous as it is convenient, easy and time saving. It also provides platform to share experiences and provide feedback. These reviews help to know whether the product is good or bad and to make decisions by potential customers. Sentiment analysis is the method by which information can be fetched from these reviews, analyzed and categorized as positive, negative or neutral. Main purpose of Sentiment analysis is to train computers to be able to understand, recognize and generate emotions. The other names for sentiment analysis are Opinion Mining, Opinion Extraction, Sentiment Mining and Subjective Analysis. Different challenges in this process, methods and levels on which sentiment analysis can be performed are discussed in this paper.

**Key words:** Reviews, Polarity, Sentiment Analysis, Sentence Level, Document Level, Feature Level

## I. INTRODUCTION

Human life is filled with emotions and expressions. They play a critical and vital role in how we think, behave and act. With the exponential increase in the internet usage, people prefer to express and share their views on internet about different topics ranging from products to hospitals to education and different services. Due to ever increasing existence of these reviews, opinions, feedbacks and suggestions on the web, it becomes necessary to analyze, organize and store these opinions for better decision making by future users. This new research domain is called sentiment analysis. It is one of the hottest research areas in computer science nowadays. Many researchers are focusing on this field and several techniques and tools have been developed over the years. Sentiment classification, subjectivity classification, aspect based sentiment analysis and cross domain sentiment analysis are some of the research directions.

## II. MAJOR CHALLENGES IN SENTIMENT ANALYSIS

Language is the most wonderful, dynamic and mysterious phenomenon in the universe [2]. Language and its structure is the primary challenge. There are several issues and challenges in the field of sentiment analysis that need to be addressed to get the genuine result [1]. Few of them are:

### A. Word Sense Disambiguation

Often encountered problem is Word Sense Disambiguation WSD [3, 4]. Correct meaning of word based on the context needs to be extracted as word can have different meanings for different domains. For example “big size” can be negative opinion for mobile phones but positive for hotels.

### B. Comparisons

To determine the polarity for comparative sentences can be a challenge. For example “Room service of hotel X is better than hotel Y”. This review has positive word 'better' but the author's preferred object is not easy to determine which is the key piece of information in a comparative reviews.

### C. Negations

Feedbacks having ‘not’, ‘never’, ‘no’ etc need to be handled very carefully otherwise it can give completely wrong results. For example “It is sure that this phone will not break easily”. This review shows positive polarity but presence of negation changes the effect completely.

### D. Intensity

Depending upon the intensity of opinion (mild or strong), to obtain result as “highly positive” or “highly negative” can also be challenging. It is called as degree of polarity.

### E. Sarcasm

People can express opinion in express ways. Another interesting challenge can be of identification of sarcasm and to analyze emotions expressed in text at a more fine-grained level.

### III. DATA SOURCE

The rapid proliferation of Internet services has led to ever increasing large volumes of e-commerce and online reviews/opinions. User opinions have a direct impact on the improvement of the quality of services. There are different sources on internet which can help to understand the reception levels of products and services provided to customers.

#### A. Blogs

Nowadays blog writing has become very common. People love to share their opinions about any particular product or general issues or their other interests. This information play vital role and can be extracted and used for the study of sentiment analysis.

#### B. Review Sites

Customers always prefer to know about the experience of others for a particular product before buying it. A large number of user-generated reviews are available on the Internet which can be referred by future users. The reviewers data used in most of the sentiment classification studies are collected from the e-commerce websites like www.amazon.com (product reviews).

#### C. Datasets

Sentiment analysis is done mostly on movie reviews. The dataset can contain different class of product reviews or multi domain reviews extracted from Amazon.com including Books, DVDs, Electronics and Kitchen appliances.

#### D. Pre-Processing of Text

It is very necessary to pre-process the text before classifying it positive or negative. It is cleaning and preparing text for classification. Some of feedbacks may contain unimportant parts called noise such as advertisements, emoticons, scripts etc. This preprocessing can also be done for special symbols, numbers, commas, separators etc. as they do not affect the orientation of text. Some question words like what, where, why also do not help to find the polarity. URLs should also be removed. Filtering is one more task for preprocessing. In any opinion if word like ‘ammmmmazing’ with repeated letters appear, it shows the intensity of expression but it is of no use its orientation. In addition, on words level, many words in the text do not have an impact on the general orientation of it. Keeping those words makes the dimensionality of the problem high and hence the classification more difficult since each word in the text is treated as one dimension. If noise in the text is reduced properly, it should help to improve the performance of the classifier and speed up the classification process. The whole process involves several steps: online text cleaning, white space removal, expanding abbreviation, stemming, stop words removal, negation handling and finally feature selection.

### IV. POLARITY DETECTION AND CLASSIFICATION

Polarity is calculated on the basis of majority of opinion words present in the review document. Analysis on each word is done to categorize that word whether positive or negative. If the number of positive words in review is more, then the polarity of document is positive. If number of negative words is more, then the polarity is negative otherwise with equal number of positive and negative words, the document falls under the neutral polarity. The text can be classified into one of the three categories as:

#### A. Positive Opinion

The system gives the positive opinion to the document if the number of positive opinion words is more than the number of negative opinion words in the document.

For example review for a hotel like “The Location of that hotel is very nice but to travel that distance is very *tiring*. Room service is also awesome. Overall it offers luxury amenities.”

This review document falls under positive polarity because the count of positive words is more than the negative words.

#### B. Negative Opinion

The system gives the negative opinion to the document if the number of negative opinion words is more than the number of positive opinion words in the document.

For example document like “In this city mostly hotels are in bad condition. Efforts taken by managers and owners are good but they have been unable to get the customers.”

This document shows negative polarity because here the count of negative opinion words is more than the positive opinion words.

#### C. Neutral Opinion

The system gives the neutral opinion to the document if the number of positive opinion words is equal to the number of negative opinion words in the document.

“I like commercial movies but I get *bored* easily. I watch movies when I am free”

This document shows neutral polarity because here the count of negative opinion words is equal to the positive opinion words.

Degree of polarity can also be calculated for some cases depending on the results required. Degree of polarity can be like highly positive or extremely negative as per the score calculated.

## V. DIFFERENT LEVELS OF SENTIMENT ANALYSIS

Sentiment analysis can be done through various methods and techniques. Some online tools are also available. Depending upon the granularities required, three different levels on which sentiment analysis can be performed are:

### A. Document Level Sentiment Analysis

This is the simpler form of classification. The whole document of opinionated text is considered as basic unit of information. This classification can be done only if the document is having opinion about single object only (like movie, book or hotel). This approach is not suitable if document contains opinions about more than one object as in forums and blogs. Classification for the whole document is done as positive or negative. Irrelevant sentences or words need to be removed before processing.

A lot of work [5, 6, 9] has been done on document based sentiment analysis. There are two methods to do classification.

- Supervised machine learning approach
- Unsupervised machine learning approach

In supervised machine learning approach there is finite set of classes for classification. Training dataset is also prepared and available. Given the training data, the system classifies the document by using one of the common classification algorithms such as K Nearest Neighbours, Support Vector Machine, Naïve Bayes and Maximum Entropy etc. Document based classification has been done by authors of [9] for news comments using different supervised machine learning approaches. Different approaches can be combined also for effective results. The Naive Bayes and Neural Network classifier are combined by authors of [10] for classifying movie reviews. They proved that accuracy of sentiment analysis is enhanced up to 80.65% by combining these two methods.

In different researches done by various authors [5, 6, 8] data is analysed using unsupervised machine learning approach. In unsupervised approach, Sentiment Orientation (SO) of opinion words in document is considered. If the SO of these words is positive then the document is classified as positive otherwise negative. In the most prominent work done by author Turney in 2002 [6], two words “Poor” and “Excellent” have been used. The semantic orientation determines whether meaning of opinion is closer to positive word “Excellent” or negative word “Poor”. Point Wise Mutual information method is used to calculate the semantic orientation. Lexicon based method has been used to perform sentiment classification [7]. The unsupervised dictionary based technique (WordNet) is used by (Richa Sharma *et al*) [5] to determine the polarity of the movie reviews at document level. In this paper Seed list contains opinion words along with their polarity. The overview of this system is:

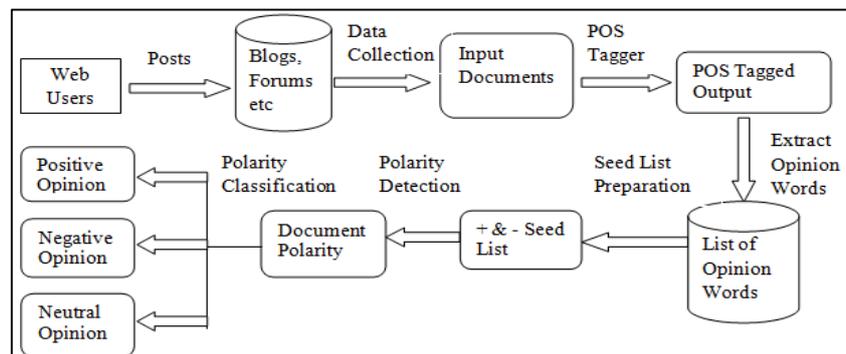


Fig. 1: Document based Sentiment Orientation System

### B. Sentence Level Sentiment Analysis

Sentence level sentiment analysis is the most fine-grained analysis of the document. In this, polarity is calculated for each sentence as each sentence is considered as separate unit and each sentence can have different opinion. Sentence level sentiment analysis has two tasks:

#### 1) Subjectivity Classification

A sentence can be either subjective sentence or objective sentence. Objective sentence contains the facts. It has no judgement or opinion about the object or entity while subjective sentence has opinions (e.g.), “India’s economy is heavily dependent on tourism and IT industry. It is an excellent place to live in.” The first sentence is a factual one and does not convey any sentiment towards India. Hence this should not play any role in deciding on the polarity of the review and should be filtered out [3]. The advantage of sentence level analysis lies in the subjectivity/ objectivity classification.

#### 2) Sentiment Classification

Sentence can be classified as positive, negative or neutral depending upon the opinion words present in it.

A number of researches [12, 13, 14] focus on finding how to classify the text effectively. The same methods of document level classification can be applied to the sentence level classification problem. A number of different methods are discussed and compared under supervised machine learning approach by authors of [11]. The research done by authors in [12] defines pure, short and no unrealistic sentences in addition to the subjective sentences. They observe that first and last lines of a review are often indicative of review polarity. An in depth analysis on different sentence types has been performed by them. A rule based domain independent sentiment analysis and lexical approach for classification of objective and subjective sentences has been proposed by Aurangzeb Khan and Baharum Baharudin in [4]. The semantic score of subjective sentence is extracted from SentiwordNet lexical resource.

### C. Feature Level Sentiment Analysis

Feature engineering is an extremely basic and essential task for Sentiment Analysis [3]. The basic logic in Feature Level sentiment analysis is to identify the piece of text as a feature of some product. For example "Memory is sufficient in this phone". In this review 'memory' is product feature (noun) and 'sufficient' is opinion word (adjective).

Minqing Hu and Bing Liu's work [16] is the most pioneering in this field. They presented different techniques based on data mining and NLP methods like frequent features, compactness pruning, P-support pruning and infrequent feature identification. Experiments have been conducted using these techniques for various electronic products. These techniques further have been outperformed by frequent pattern mining algorithm called H-mine proposed by [17]. They state that using Apriori algorithm leads to increase the execution time while dealing with large databases. Various techniques and tools have been proposed by different researchers to enhance the performance and precision of the system.

A new approach is proposed by authors of [18] which uses feature oriented appraisal words lexicon. It is fine grained approach in which review categorization is based on attitude and polarity of the adjectival words for the frequent features of the product.

## VI. CONCLUSIONS

Sentiment Analysis is an emerging field of data mining. It is becoming more essential as more people preferably like to buy online and give feedback, reviews and comments on the products. This paper discusses about the basic steps to be performed before sentiment analysis, different resources for the reviews, an overview of Sentiment Analysis and its various level on which this analysis can be conducted in form of Sentence, Document and Feature level sentiment analysis. Various challenges are also discussed that make sentiment analysis a difficult task.

## REFERENCES

- [1] Seema Kolkur, Gayatari Dantal, Reena Mahe, "Study of Different Levels for Sentiment Analysis" (2015) IJCET
- [2] Jaganadh G Opinion, Mining and Sentiment Analysis, (2012), CSI Communications
- [3] S.ChandraKala and C. Sindhu, " Opinion Mining and Sentiment Classification: A Survey", ICTACT Journal on Soft Computing, Vol- 03, ISSUE: 01, ISSN: 2229-6956, (2012)
- [4] Aurangzeb Khan, Baharum Baharudin, " Sentiment Classification by Sentence Level Semantic Orientation using SentiWordNet from Online Reviews and Blogs", Int. J Comp Sci. Emerging Tech, Vol-2 No 4, (2011)
- [5] Richa Sharma, Shweta Nigam and Rekha Jain, "Opinion Mining Of Movie Reviews At Document Level", International Journal on Information Theory (IJIT), Vol.3, No.3, (2014)
- [6] Peter D. Turney, "Thumbs Up or Thumbs Down ? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp. 417-424, (2002)
- [7] M. Taboada, J. Brooke, M. Tofiloski, K. Voll , and M. Stede, "Lexicon-based methods for sentiment analysis," Association for Computational Linguistics, vol. 37, (2011)
- [8] Gautama Kumar, Pawan kumar Goel, Sanjeev kumar Chauhan, Anand kumar Pandey, "Opinion mining and summarization for customer reviews", International Journal of Engineering Science and Technology (IJEST), Vol. 4 No.08, and ISSN: 0975-5462, (2012)
- [9] Yan Zhao, Suyu Dong and Leixiao Li, " Sentiment Analysis on News Comments Based on Supervised Learning Method", International Journal of Multimedia and Ubiquitous Engineering, Vol.9, No.7 pp.333-346, (2014)
- [10] Lina L. Dhande and Dr. Prof. Girish K. Patnaik, "Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 4, ISSN 2278-6856, (2014)
- [11] S Padmaja and Prof. S Sameen Fatima, " Opinion Mining and Sentiment Analysis –An Assessment of Peoples' Belief: A Survey", International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC), Vol.4, No.1, (2013)
- [12] V. S. Jagtap, Karishma Pawar,"Analysis of different approaches to Sentence-Level Sentiment Classification", International Journal of Scientific Engineering and Technology, PP: 164-170, (2013)
- [13] Raisa Varghese, Jayasree M, "A Survey on Sentiment Analysis and Opinion Mining", International Journal of Research in Engineering and Technology (IJRET), eISSN: 2319-1163 pISSN: 2321-7308, (2013)
- [14] Gizem Gezici, Berrin Yanikoglu, Dilek Tapucu, and Yucel Saygin, "New Features for Sentiment Analysis: Do Sentences Matter?" First International Workshop on Sentiment Discovery from Affective Data (SDAD), (2012)
- [15] Padmapani P. Tribhuvan, S.G. Bhirud, Amrapali P. Tribhuvan, "A Peer Review of Feature Based Opinion Mining and Summarization", International Journal of Computer Science and Information Technologies, Vol. 5 (1), ISSN: 0975-9646, (2014)
- [16] Minqing Hu and Bing Liu, "Mining and Summarizing Customer Review", ACM, (2004)
- [17] Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi,"A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews" International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, ISSN: 1694-0814, (2012)
- [18] D D Chaudhari, R A Deshmukh, A B Bagwan, P K, (2013)