

---

## Sentiment Analysis Based on Dictionary Approach

Reshma Bhonde<sup>1</sup>, Binita Bhagwat<sup>2</sup>, Sayali Ingulkar<sup>3</sup>, Apeksha Pande<sup>4</sup>

Department of Computer Engineering SKN-Sinhgad Institute of Technology and Science, Lonavala,  
Maharashtra, India

<sup>1</sup>reshma.bhonde@gmail.com, <sup>2</sup>binita294@gmail.com, <sup>3</sup>sayaliingulkar@gmail.com,  
<sup>4</sup>apekshapande65@gmail.com

---

**Abstract:** *Sentiment analysis has become upcoming topic in recent years, and sentiment dictionaries are essential for research in this field. Consumers regularly face the trade-off in purchase decisions so nowadays if one wants to buy a consumer product one prefer user reviews and discussion in public forums on web about the product. Many consumers use reviews posted by other consumers before making their purchase decisions. People have a tendency to express their opinion on various entities. As a result opinion mining has gained importance. Sentiment Analysis deals with evaluating whether this expressed opinion about the entity has a positive or a negative orientation. Consumers need to decide what subset of available information to use. Graphical analysis is a better option for this decision as they provide a quick indication of a review. However there are cases where such graphical representation are not available or detailed enough. Sentiment analysis text analysis techniques that automatically detect the polarity of text can help in these situations with more refined analysis.*

**Keywords:** *Opinion mining, sentiment analysis, natural language processing (LNP), text mining.*

---

### 1. INTRODUCTION

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Generally speaking, Sentiment analysis with respect to the subject of speaker or an author's point of view or a document to determine the overall functionality. Sentiment Analysis uses various classification techniques to identify the tone of a given piece of text. It indicates whether the text is positive, negative or neutral. This analysis can be aggregated over large sets of data and the resulting information can be helpful in different contexts.

Nowadays, social media has become a platform for people to convey their voice to the public. The Internet has rapidly advanced from a static to an interactive medium. Today's users cannot only obtain information but also actively generate content. News reports, BBS, forums, blogs, etc. are the main sources of public opinion information. The text contains both cases and opinion which could be extracted using natural language processing to get some opinionated views. Sentiment analysis is also called as opinion mining. Sentiment analysis not only helps in allowing the user to get more and relevant information about different products and services on a mouse click, but also helps in arriving at a more informed decision. The analysis of sentiments may be document based where the sentiment in the entire document is summarized as positive, negative or objective. It can be sentence based where individual sentences, bearing sentiments, in the text are classified. SA can be phrase based where the phrases in a sentence are classified according to polarity. In fact, to identify the emotion analysis task views expressed in a text is positive or negative weather. Natural language processing (NLP) computer science, Artificial intelligence, and computers and human (natural) concerned with interactions between languages is an area of Linguistics. For instance, ina product review, it identifies features of the product that have been commented on by the reviewer and determines whether the comments are positive, negative or neutral. For example, in the sentence, "The life of the battery of this mobile is too compressed", the opinion is on "life of the battery" of the mobile object (target) and the opinion is negative. Many day to day life applications require this level of detailed analysis because in order to make product upgrade one needs to know what components and/or features of the product are liked and disliked by consumers. Such information has not come across by sentiment and subjectivity classification.

## 2. OPINION MINING AND SENTIMENT ANALYSIS

The concept of opinion mining is nothing but data mining. The sentiment analysis problem is met using some of the techniques using natural language processing technique, proximity method etc. People give their opinions on anything, e.g. one buy a product and then express their views on products features in various forums. The term object is used for the entity on which comments have been given. A notable approach in uses a sentence level sentiment analysis. The word level feature abstraction is done using Naive Bayesian Classifier. For classifying and analyzing of the sentiment from the reviews, machine learning and lexical contextual information are used. The naive bayes approach is used to interpret each sentence as positive and negative on the bases of useful word level feature. SVM classifier is trained on the interpreted sentences for the positive and negative classification. Contextual data is used to calculate the polarity of sentence and mark it as either negative or positive.

*Definition (Opinion):* An opinion is a quadruple, (g, s, h, and t), where g is the opinion (or sentiment) target, s is the sentiment about the target, h is the opinion holder and t is the time when the opinion was expressed.

*Definition (opinion holder):* opinion on the person the holder of his/her opinion or negative opinion (or sentiment) on an object.

### 2.1. Naïve Bayes

The naïve bayes algorithm uses Bayes theorem. The formula  $P(C|F)$  states the conditional probability of C given F, where C is a class label and F a feature.

$$P(C|F) = \frac{P(C)P(F|C)}{P(F)}$$

It allows calculating unknown conditional probability from a known conditional probability together with the prior probabilities. It is assumed that the presence of a feature is unrelated to the presence of any other feature. The major advantage of naïve bayes models is the fact that a relatively small training set is sufficient to train the model. It is a good model to use as reference for testing the quality of other models. The naïve bayes classifier has been applied in quite a lot of papers about sentiment analysis.

### 2.2. Machine Learning

Sentiment analysis mostly belongs to supervised classification in general and text classification techniques in particular and to this machine learning approach is applicable. Thus, it is called "the supervised learning". Two sets of documents are required for machine learning based classification: training set and test set. A training set to learn the difference signs documents used by automatic classifier, and a test set to validate the performance of the automatic classifier is used for machine learning techniques to classify a number of review adopted. Naive Bayes (NB), maximum entropy (me), machine learning techniques like and support vector machine (SVM) has achieved great success in text classification. Other noted machine learning ways the subsequent natural language processing and Qiang Ye, 2009. The basic idea combined words and categories of prospects, given test document categories using prospects. Such model is the notion of independence is part of the word naïve simplicity of perception Naïve Bayes Classifier of the computation makes far more efficient. Support vector machines (SVM), a discriminative classifier is considered the best text classification method (Rui Xia, 2011; Ziqiong, 2011; Songho tan, 2008 and Rudy Prabowo, 2009). . The support vector machine is a statistical classification method proposed by Vapnik. Computational learning theory based on minimum principle of structural risk, SVM training data points to the surface a decision to separate the two sections and support vector training set is effective elements in selected decisions based on many of the variants. SVM spirit in which multiple class classification (Kaiquan Xu, 2011) is used for has been developed.

Variety of feature sets and classification algorithms to synthesize a more precise classification process by integrating emotion classification for a comparative study of the effectiveness of education and technology. in his work the spirit of two types of feature sets are designed for classification, Namely the part and text-to-speech feature sets-based feature-based classification algorithms then, three text sets, namely Bayes, maximum entropy and support vector machine classifiers feature sets, naïve base-

## Sentiment Analysis Based on Dictionary Approach

as for each of the classification score to predict. three types of ensemble methods, namely fixed combinations, and weighted metaclassifier combination, are rated for three ensemble strategies IE feature set ensemble classification algorithms of ensemble, Facility, and both ensemble sets and classification algorithms.

### 2.3. Dictionary-Based Approach

Using a dictionary approach to compile sentiment words is an obvious approach because most dictionaries (e.g., WordNet (Miller, 1990)) list synonyms and antonyms for each word. Thus, a simple technique in this approach is to use a few seed sentiment words to bootstrap based on the synonym and antonym structure of a dictionary. Specifically, this method works as follows: A small set of sentiment words (seeds) with known positive or negative orientations is first collected manually, which is very easy. The algorithm then grows this set by searching in the WordNet or another online dictionary for their synonyms and antonyms. The latterly found words are added to the seed list. The next iteration begins. The iterative process ends when no more new words can be found. After the process completes, a manual inspection step was used to clean up the list.

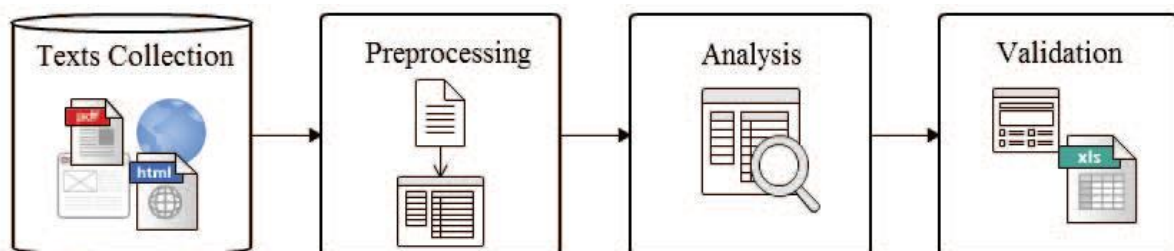
### 2.4. Word level Sentiment Analysis

In the word based approach the criterion for selection a tweet to automatic classification is the presence of words that express sentiment such as good, bad, excellent or terrible. From these words it is possible to infer the sentiment present in the text. This words are used to determine the sentiment (positive and negative) according to the application. This is the simplest form of sentiment analysis and it is assumed that the word contains an opinion on one main object expressed by the author of the document.

### 2.5. Sentence Level Sentiment Analysis

This is the simplest form of sentiment analysis and it is assumed that the document contains an opinion on one main object expressed by the author of the document. Various papers have been written on this topic. The supervised way assumes that there is a finite set of classes into which the document should be classified and training data is available for every class. The simplest case is when there are two classes: positive and negative. Simple expansions can also add a neutral class or have some discrete numeric scale into which the record should be placed (like the five-star system used by Amazon). More progressive representations utilize TFIDF, POS (Part of Speech) information.

## 3. TEXT MINING PROCESS



**Fig1.** The Text Mining Process

Text mining corresponds to a set of techniques used to extract pattern or trends implicit in textual databases. Text mining process can be divided into four steps text collection; preprocessing; analysis and validation.

### 3.1. Text Collection

Capturing the text is the first step of the process and aims at generating the data. These database can be static or dynamic. Static database remains the same throughout the process while dynamic database can be updated at every instant of time.

### 3.2. Pre-processing

Pre-processing starts the text preparation into a more structured representation.

- 1) *Tokenization*: Tokenization is used to identify all words in a given text.
- 2) *Data Filtering*: People use a lot of casual language on twitter. For example, 'happy' is used in the form of 'haaaaaappy'. Though this implies the same word 'happy', the classifiers consider these as

two different words. To improve this and make words more similar to generic words, such sets of repeated letters are replaced by two occurrences. Thus haaaaappy would be replaced by haappy.

- 3) *Stop Word Removal*: Is used to eliminate that words that occurs frequently such as article, prepositions, conjunction and adverbs. These stop words depends on language of the text in questions. For example, words like the, and, before, while, and so on do not contribute to the sentiment.
- 4) *Stemming*: In information retrieval, stemming is the process of reducing a word to its root form. For example, walking, walker, walked all these words are derived from the root word walk. Hence, the stemmed form of all the above words is walk.

### 3.3. Analysis

The analysis step is usually considered the core of text mining, because this is when some type of useful, nontrivial knowledge is extracted from the text.

### 3.4. Validation

In order to validate the analysis are performed. It is necessary to employ quantitative measures and qualitative measures. After such a validation it may be necessary to return to one or more of the previous step so as to perform modification and try alternatives.

## 4. SENTIMENT CLASSIFICATION PERFORMANCE METRICS

Generally, the performance of sentiment classification is evaluated by using four indexes: Accuracy with Precision plus Recall and F1-score. This is the normal option to compute these indexes which are depend on the confusion matrix

The equations show the indexes:

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{Accuracy} = \frac{tp + tn}{tp + fn + tn + fp}$$

$$\text{F-measures} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

The accuracy of prediction predicts all against all instances of true examples. Instances of a 100% accuracy to predict exactly the same as the actual examples are meant to predict all positive instances against true. Precision predicts positive instances. Remember that part of the true positive prediction turned positive in all instances is against real F1 recall and precision of a harmonic.

## 5. CONCLUSION

In this paper we have discussed about the sentiment analysis and opinion mining in detail. Sentiment Analysis deals with evaluating whether this expressed opinion about the entity has a positive or a negative orientation. We observed that opinion mining helps the consumer to take the right decision after particular thing is analyzed.

## REFERENCES

- [1] Automatic sentiment Analysis of Twitter Message by Ana C.E.S.Lima and Lean draN.Castro @ 2012 IEEE.
- [2] Is Worth a Thousand Words: How People Attach Sentiment to Emoticons and Words in Tweets978-0-7695-5137-1/13 \$26.00 © 2013 IEEE DOI 10.1109/SocialCom.2013.54

- [3] Development of a Novel Algorithm for Sentiment Analysis Based on Adverb-Adjective-Noun Combinations 978-1-4577-0748-3/12/\$26.00 © 2012 IEEE.
- [4] Star Ratings versus Sentiment Analysis - A Comparison of Explicit and Implicit Measures of Opinions 978-1-4799-2504-9/14 \$31.00 © 2014 IEEE DOI10.1109/HICSS.2014.106
- [5] Sentiment Analysis and Opinion Mining April 22, 2012 Bing Liuliub@cs.uic.edu
- [6] Study on the Method of Identifying Opinion Leaders Based on Online Customer Reviews.
- [7] SA literature survey 2012
- [8] SA literature survey 2012
- [9] The How, When and Why of Sentiment Analysis IJCTA | July-August 2013 Available online@www.ijcta.com
- [10] A Survey on Sentiment Analysis and Opinion Mining IJRET | Nov-2013, Available @ <http://www.ijret.org>
- [11] Twitter Sentiment Analysis CMPS 242 Project Report Shachi H Kumar University of California Santa Cruz Computer Science shachihkumar@soe.ucsc.edu