

## A Methodological Framework for Opinion Mining

Angelpreethi<sup>1\*</sup>, P. kiruthika<sup>2</sup>, S. BrittoRameshKumar<sup>3</sup>

<sup>1</sup>Department of Computer Science, St. Joseph's College, Trichy-2

<sup>2</sup>Department of Computer Science, Urumu Dhanalakshmi College,

<sup>3</sup>Department of Computer Science, St. Joseph's College, Trichy-2.

\*Corresponding Author: [angelpreethi.mca@gmail.com](mailto:angelpreethi.mca@gmail.com), *mob: 9578222116*

Available online at: [www.ijcseonline.org](http://www.ijcseonline.org)

**Abstract** -Our day to day life has always influenced by what people think. Opinion and ideas always precious our own opinions. Sentiment Analysis is machine learning approach in which machine analyzes and classifies the human's sentiments, emotions, and opinions about some topic which are expressed in the form of either text or speech. Sentiment Analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. Till now, there are few different problems predominating in this research community, namely, sentiment classification, feature based classification and handling negations. In real world, public or consumer opinions about some product or brand are very important for its sales. Hence sentiment analysis is a very important research area for real life applications i.e. decision making. Hence this paper aims to cover different algorithms of opinion mining.

**Keywords**—Sentiment Analysis, Opinion Mining, Machine Learning, Classification, Sentiment Polarity

### I. INTRODUCTION

Every field of life is growing quickly and generating data. As a result huge volumes of data produced in the field of science, engineering, medical, marketing, finance, demographic etc. For the classification, analysis and summarizing data requires automated systems. Visualization, statistics, machine learning are the research areas involved to develop automated systems. In a democracy people express their views and opinions in several ways. One way is through web. People also want to see the opinion of other persons about a product to analyze it before buying the particular product. Opinion is a vital part in our day to day life. Sentiment analysis and Opinion mining is a type of natural language processing for track the mood of the public about a particular product or topic. For example, in marketing it helps judging the sensation of new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features. There are many challenges in sentiment analysis. The first is that an opinion word that is considered to be positive in one state may be considered negative in another state. A second challenge is that people don't express opinions in a same way.

Most reviews will have both positive and negative interpretations, which somewhat manageable by analyzing sentences one at a time. However in more unofficial medium like twitter or blogs, the more likely people are to combine diverse opinions in the same sentence which is easy for human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty to

understand what someone's thought based on a short piece of text because it lacks context. The sentiment analysis problem can be solved to a satisfactory level by manual training. But a fully automated system for sentiment analysis which needs no manual intervention has not been introduced yet.

### II. DIFFERENT LEVELS OF SENTIMENT ANALYSIS

#### A. Phrase Level Sentiment Analysis

This classification is much more pinpointed approach to opinion mining. The phrases that contain opinion are found out and a phrase level classification is done.

#### B. Sentence Level Sentiment Analysis

The sentence level sentiment analysis, the polarity of each sentence is calculated. The same document level classification methods can be applied to the sentence level classification problem.

#### C. Document Level Sentiment Analysis

In this document level classification, a single review about a single topic is considered.

### III. DIFFERENT CLASSES OF SENTIMENT ANALYSIS

Sentiments can be classified into three classes. (Ex) positive, negative and neutral sentiments.

- a. **Positive Sentiments:** These are the good words about the goal in reflection. If the positive sentiments are increased, it is referred to be good. In

case of product reviews, if the positive reviews about the product are more, it is bought by many customers.

- b. Negative Sentiments:** These are the bad words about the goal in reflection. If the negative sentiments are increased, it is discarded from the favorite list. In case of product reviews, if the negative reviews about the product are more, no one propose to buy it.
- c. Neutral Sentiments:** These are neither good nor bad words about the goal. Hence it is neither preferred nor neglected.

#### IV. CLASSIFICATION ALGORITHMS

##### a. Naïve Bayes Classifier:

The Naive Bayes classifier is a simple probabilistic classifier which is based on Bayes theorem with strong and naïve independence assumptions. It is one of the most basic text classification techniques with various applications in email spam detection, personal email sorting, document categorization, language detection and sentiment detection.

##### b. Support Vector Machine:

SVM is generally used for text categorization. SVM gives best results than Naive bayes algorithm in case of text categorization. The basic idea is to find the hyper plane which is represented as the vector  $w$  which separates document vector in one class from the vectors in other class. SVM also supports classification and regression in statistical learning.

##### C. Multi Layer Perceptron:

Multi-Layer perceptron is a feed forward neural network, with one or N layers among inputs and output. Feed forward means i.e uni-direction flow of data such as from input layer to output layer. This Artificial Neural Network which multilayer perceptron begin with input layer where every node means a predictor variable.

#### LITERATURE REVIEW

##### 4.1. A survey on sentiment analysis and opinion mining:

[Dudhat Ankitkumar.M, Badra R.R, Mayura kinikar]

Sentiment analysis is machine learning approach in which machine analyzes and classifies the human's sentiments, emotions, and opinions about some topic which are expressed in the form of either text or speech.

##### Pros:

- To classifies the sentence into positive, negative or neutral.

- Two main approaches are to be used from this sentiment analysis. If there are machine learning and lexicon based approach.

##### Cons:

- The lexicon method cannot find the opinion words with domain and context specification orientations
- SVM gives good accuracy compare to other classification techniques, but still it has some limitations.

##### 4.2. Sentiment analysis and opinion mining: A survey:[Vinodhini.G, Chandrasekaran.RM]

A few recent studies in this field explained the use of neural networks in sentiment classification. Zhu Jian (2010) proposed an individual model based on Artificial neural networks to divide the movie review corpus into positive , negative and fuzzy tone which is based on the advanced recursive least squares back propagation training algorithm.

##### Pros:

- Two types of feature sets are designed for sentiment classification, namely the part-of-speech based feature sets and the word-relation based feature sets.
- The machine learning approach is more accurate but requires a significant amount of time to train the model.

##### Cons:

- Sentiment analysis can be applied for new applications.
- A lot of problems in this field of study remain unsolved.
- The main aspects exist in use of other languages, dealing with negation expressions, handling of implicit product features.

##### 4.3. A survey on sentiment analysis on twitter data using different techniques: [ Bholane Savita Dattu, Deilpali V. Gore]

Twitter sentiment analysis is an important research area for academic as well as business fields for decision making like for the seller to decide if the product should be produced in a large quantity as per the buyers feedback and for the students to decide if the study material to be referred or not. So to do the sentiment analysis accurately without misconceptions the event topic model is very useful.

##### Pros:

- Machine learning methods minimize the structural risks.
- The problem of unbalanced dataset in sentiment classification is solved efficiently and appropriately.

##### Cons:

- The sentiment analysis tools Twitter Sentiment and SentiStrength whose accuracy is less as compared to other sentiment analysis techniques.

- Supervised methods require excessive -quantity of labeled training dataset which are very expensive.

#### 4.4 A survey on sentiment analysis and opinion mining:[

Raisa Varghase, Jayasree M]

The naive bayes approach is used to annotate each sentence as positive and negative on the bases of useful word level feature. SVM classifier is trained on the annotated sentences for the positive and negative classification. Contextual information is used to calculate the polarity of sentence and mark it as either negative or positive.

##### Pros:

- Cross domain sentiment analysis is introduced to reduce the manual effort in training the machine using labeled data.
- Sentiment sensitivity is achieved in the thesaurus by incorporating document level sentiment labels in the context vectors used as the basis for measuring the distributional similarity between words.

##### Cons:

- Still exist in the field of machine learning and some of them are Named entity Recognition, Co reference Resolution, domain dependency.
- These problems have to be tackled separately and those solutions can be used to improve the methods to do sentiment analysis.

#### V. SENTIMENT ANALYSIS FRAMEWORK:

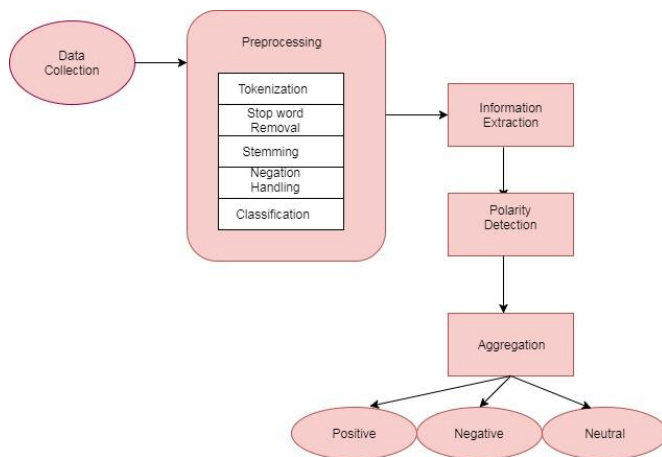


Figure 1. Sentiment Analysis Framework

Data Collection: The first step is data collection; data is collected from networking site like Blogs or may be collect from e-commerce site. This collected data is noisy and cannot give good information so analysis can be done on this data. Natural Language Processing or other text mining can be done on this data to take out information.

Data Preparation: The second step is data preparation in this step noisy data can be cleaned and prepare for sentiment analysis.

Information Extraction: The third step is sentiment detection analysis opinion in extract for data this opinion may be a facts, attitude or entities of review's that posted by user online about the product.

Polarity Detection: The fourth step is text is classified according to negative, positive, neutral, better, amazing, bad, worst, good, and wonderful. The classification can be done on only my point of views.

Aggregation: The final step is represent the output of analysis data, the output is show in a graphical layout it may be in line chart, pie chat or in any other graphical demonstration it include time average a point of classification that use in analysis on data for sentiment analysis.

#### VI. CONCLUSION

Sentiment Analysis problem is a machine learning problem that has been a research interest for recent years. Through this literature survey, the relevant works done to solve this problem could be studied. Naive Bayes classifier is insensitive to unbalanced data which give more accurate results. If it will found on to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification performance.

#### REFERENCES

- Dudhat Ankitkumar.M, Badra R.R, Mayura kinikar, "A Survey on sentiment analysis and opinion mining", International Journal of Innovative Research in Computer and Communication Engineering, vol2, issue.11, Nov 2014.
- Vinodhini.G, Chandrasekaran.RM, "Sentiment analysis and opinion mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, vol2, issue.6, June 2012.
- Bholane Savita Dattu, Deilpali V. Gore , "A Survey on sentiment analysis on twitter data using different techniques", International Journal of Computer Science and Information Technologies, vol 6, issue.6, June 2015.
- Raisa Varghase, Jayasree M, "A Survey on sentiment analysis and opinion mining", International Journal of Research in Engineering and Technology, vol2, issue.11, Nov 2013.
- Anu Sharma, Savleen kaur, "Techniques for sentiment analysis survey", International Journal of Computer Technology and Application, Nov 2016. [6] M.Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04), pp. 168–177, August 2004.
- B. Liu, M. Hu, and J. Cheng, "Opinion observer: analyzing and comparing opinions on the web," in *Proceedings of the International World Wide Web Conference Committee (IW3C2 '05)*, Chiba, Japan, pp.10–14, May 2005.
- Krisztian Balog, Maarten de Rijke. 2006. Decomposing Bloggers' Moods - Towards a Time Series Analysis of Moods in the Blogosphere.

- [9] Khairullah Khan, Baharum B. Baharudin, Aurangzeb Khan, Fazal-e-Malik, "Mining Opinion from Text Documents: A Survey", 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009.
- [10] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," presented at the Proceedings of the 12th international conference on World Wide Web, Budapest, Hungary, 2003.
- [11] B. Liu. *Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing*, Second Edition, (editors: N. Indurkha and F. J. Damerau), 2010.
- [12]. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2(1-2), pp. 1–135, 2008.
- [13] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Comput. Linguist.*, vol. 37, pp. 9-27, 2011.
- [14] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques," presented at the Proceedings of the Third IEEE International Conference on Data Mining, 2003.
- [14] N. Jakob and I. Gurevych, "Extracting opinion targets in a single- and cross-domain setting with conditional random fields," presented at the Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, Massachusetts, 2010.
- [15] W. Jin, H. H. Ho, and R. K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.