

# Sentiment Analysis: Classification and Approaches

Sowmya. B<sup>1</sup> Ramya. J<sup>2</sup>

<sup>1,2</sup>Assistant Professor

<sup>1,2</sup>Department of Computer Science Engineering

<sup>1,2</sup>SaIT Bangalore, India

**Abstract**— The rapid growth of textual information on web has brought radical change in human life. In web people share their views and opinions about products or services they aware of. This leads to a large collection of views and opinions in the form of texts, which needs to be analyzed and categorized to know the efficiency of the product and service. Opinions are expressions that describes human's sentiment or feeling towards the product or service. This information is very useful for businesses, governments and individuals. It is very difficult and time consuming to analyze this bulk amount of user generated data, so there is a need of intelligent system which automatically analyze such huge data and classify them as positive, negative and neutral. Sentiment analysis is the process of extracting opinions, emotions and attitude from text, speech and database sources through Natural Language Processing. This survey is a summary of the sentimental level of classification that assigns polarity to opinionated text, widely used algorithms, possible challenges and application areas of sentiment classification.

**Keywords:** Sentiment Analysis: Classification and Approaches

## I. INTRODUCTION

Sentiment can be described as emotions, opinions, thought, idea based on a feeling or a way of thinking about something. In Computational Linguistics, the main focus is on opinions rather than on sentiments, feelings or emotions. Sentiment Analysis is also called as subjective Analysis. It mainly focuses on the subjective part of the text to determine the outcome. Opinion mining or opinion extraction mines opinions at word, sentence and document level from different corners of the world.

Sentiment classification is the process of looking at a text and decides people's likes or dislikes the thing they are thinking about. Sentiment classification follows four steps for processing as in figure 1.

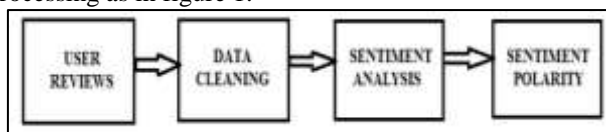


Fig. 1: process of sentiment classification

First step is to collect different reviews given by many users from social media websites. Then perform preprocessing steps (data cleaning) to clean the irrelevant data which is not needed for sentiment classification. After preprocessing linguistic features in user generated features has to be analyzed using parts of speech approach so that the required information is identified. Then sentiment classification is performed using machine learning techniques in order to determine the polarity of the text.

This paper is started with the definition of sentiment analysis and next section focus on the levels of sentiment analysis. Section 3 discusses the computational approaches

for sentiment analysis. Section 4 discusses some of the applications in this area. The last section concludes our study.

## II. LEVELS OF SENTIMENT ANALYSIS

Sentiment analysis can occur at 3 different levels:

- 1) Document Level analysis
- 2) Sentence Level analysis
- 3) Aspect/Feature Level analysis

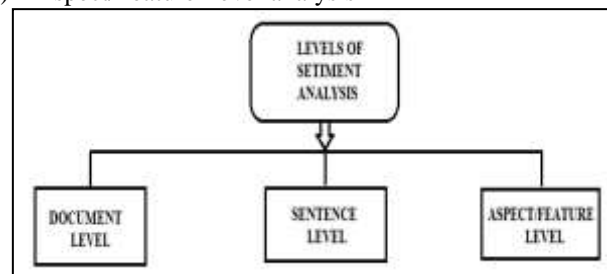


Fig. 2: Levels of sentiment analysis

### A. Document Level Analysis

The task of document level analysis is to determine the overall opinion of the document. Sentiment analysis at document level classify the review and indicate that the whole document expressed a positive or negative sentiment value. For example when a user reviewed any product the system determines an overall opinion and gives the result. Different machine learning approaches such as Naive Bayes, Maximum Entropy Classification and Support Vector Machine are used for classification of documents. Opinions that are applicable on each document is considered as a single entity and if there are multiple entities on which the result is evaluated then it is not applicable.

### B. Sentence Level Analysis

As the name implies sentiment analysis is performed on the sentences rather than performed on the whole document. This level determines that each sentence gives a positive, negative or neutral sentence level sentiment. Neutral generally means no opinion. Subjectivity classification and sentiment classification are the two main tasks that can be performed in sentence level analysis. This level of analysis is not suitable for complex structure of sentences.

### C. Aspect/Feature Level Analysis

Sentence level and document level does not discover what people like or not, so feature level sentiment analysis is more appropriate and it gives accurate analysis. Aspect based sentiment analysis is a technique that breaks down text into aspects and allocates each one a sentiment level. The results are more accurate and detailed because aspect based analysis looks very closely at the information behind a text. The aspect level classification gives the summary of multiple reviews based on the feature based opinion.

### III. COMPUTATIONAL APPROACHES FOR SENTIMENT ANALYSIS

#### A. Natural Language processing

It is the branch of computer science which is mainly focused on developing systems that allow computer to communicate with people using natural language. This technique plays an important role to get accurate sentiment analysis. opinion for document level, phrase level, sentences level and aspect level are expressed using various NLP techniques such as Bag of Words, Hidden markov model (HMM), part-of-speech, large sentiment lexicon acquisition and parsing techniques. Sentiment word dictionary is used for the acquisition large sentiment lexicons, which contains numerous amount of sentiment words with their numeric threshold value. Tagging of parts-of-speech is the most time consuming and challenging task before performing sentiment analysis of any document. As online reviews are short and contains non-grammar sentences, slangs, abbreviations and special symbols which make POS tagging more difficult.

#### B. Machine Learning Techniques

Machine learning techniques are very useful techniques for sentiment classification i.e., to categorize text as Positive, Negative or Neutral. Training dataset and testing datasets are required in this technique. The training dataset is used to learn and produce complex results. Test dataset is used for performance validation.

There are two types of machine learning techniques

- 1) Supervised Learning
- 2) Unsupervised Learning

There are various machine learning algorithms used to classify reviews such as supervised algorithms like Naive bayes, support vector machine, maximum entropy, K nearest neighbor etc., and unsupervised algorithms like Neural network, Hidden markov model, k-means etc.,

##### 1) Naive Bayes

Naive Bayes algorithm is one of the most popular and simple machine learning classification algorithm. It can be extremely fast compared to other classification algorithms and it does not require large amount of data. It is mainly use for document level classification and is based on bayes theorem for calculating probabilities and conditional probabilities of words and categories. It is optimal for problem classes with highly dependent features.

##### 2) Support Vector Machine

Support Vector Machine is a discriminative classifier and is considered as the best text classification method. Mapping of input feature vectors to higher-dimensional feature space is through some non-linear mapping. SVMs are designed on the principle of structural risk minimization. The structural risk minimization seeks a hyperplane on which user can find lowest probability of error. Computing the hyper plane to training a SVM leads to a quadratic optimization problem. Even on learning a large set of data SVMs are managed to scale better. Whenever there is a new pattern during classification SVMs can dynamically update the training patterns.

##### 3) k-Nearest Neighbor

The k-Nearest Neighbor is a simple, easy to implement algorithm which is used to solve both classification and

regression problems. KNN relies on labeled input data similar to the test document. In this method the classification of objects are based on the majority class amongst its k-nearest neighbor. KNN is a type of lazy learning in which the function is approximated locally and all computation is delayed until classification.

##### 4) Maximum Entropy

Maximum entropy is a classification technique which has proven effective in Natural Language Processing application. Maximum entropy will not make any assumptions about the features relationship, so it may perform better when conditional independence assumptions are not met.

Principle of maximum entropy is a rule in which it allows us to choose best among a number of probability distribution that express the current state of knowledge. If we does not know anything about distribution except that it belongs to certain class, then the distribution which is having largest entropy must be chosen as default.

### IV. APPLICATIONS

There are numerous applications of sentiment analysis. Sentiment analysis can be used in various fields like Movie reviews, product or service review, politics, public auctions, social media sites etc., Brand and reputation management is the most common applications of sentiment analysis. It helps many companies, marketing agencies, fashion brands, media organization and many more to know how the customer perceive their brand, product, services etc.,

Sentiment analysis allows companies to:

- 1) Track the perception of the customer about their brand
- 2) Find current trends and pattern
- 3) Point out the details about the attitude
- 4) Monitor reviews of the brands on different platforms etc.,

### V. CONCLUSION

Sentiment Analysis is also known as opinion mining is helpful in different fields mainly for expressing, calculating, identifying and classifying sentiment. In this paper, various applications and techniques for sentiment classifications are discussed. Nowadays sentiment analysis becomes the growing field of research, it plays a vital role in text mining in order to retrieve some useful information of interest.

### REFERENCES

- [1] A. Nisha, Jebaseeli, E. Kirubakaran, PhD., "A Survey on Sentiment Analysis of (Product) Reviews", International Journal of Computer Applications (0975 – 888) Volume 47– No.11, June 2012
- [2] Jalaj S. Modha, Prof & Head Gayatri S. Pandi Sandip J. Modha, "Automatic Sentiment Analysisfor Unstructured Data", International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 12, ISSN: 2277 128X, December 2013.
- [3] Raisa Varghese1, Jayasree M2, "A SURVEY ON SENTIMENT ANALYSIS AND OPINION MINING", IJRET:International Journal of Research in Engineering and Technology ISSN: 2319-1163 | ISSN: 2321-7308.
- [4] Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, "OPINION MINING AND ANALYSIS: A SURVEY",

- International Journal on Natural Language Computing (IJNLC) Vol. 2, No.3, June 2013.
- [5] Zhongwu Zhai, Bing Liu, Hua Xu and Hua Xu, "Clustering Product Features for Opinion Mining", WSDM'11, February 9–12, 2011, Hong Kong, China. Copyright 2011 ACM 978-1-4503-0493-1/11/02...\$10.00
- [6] Siddhi Patni, Avinash Wadhe, "Review Paper on Sentiment Analysis is – Big Challenge", International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 2, ISSN: 2321-7782 (Online), February 2014 .
- [7] G.Vinodhini, RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 6, ISSN: 2277 128X, June 2012.
- [8] Anderson, P., "What is Web 2.0? Ideas, technologies and implications for education", Technical report, JISC, 2007.
- [9] Mishne G. and Glance N., "Predicting movie sales from blogger sentiment", In AAAI Symposium on Computational Approaches to Analyzing Weblogs (AAAI-CAAW), 2006: 155–158.
- [10] Maria Tchalakova, Dale Gerdemann, Detmar Meurers, "Automatic Sentiment Classification Of Product Reviews Using Maximal Phrases Based Analysis", Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-HLT 2011, pages 111- 117, Portland, Oregon, USA 2011 Association for Computational Linguistics, 24 June 2011,.
- [11] Jiawen Liu, Mantosh Kumar Sarkar and GoutamChakraborty, "Feature-based Sentiment Analysis on Android App Reviews Using SAS® Text Miner and SAS® Sentiment Analysis Studio", SAS Global Forum 2013.
- [12] Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan and Claypool Publishers, p.18-19, 27-28, 44-45, 47, 90-101, May 2012.
- [13] Nitin Indurkha, Fred J. Damerau, "Handbook of Natural Language Processing", Second Edition, CRC Press, 2010.
- [14] Ronen Feldman, "Techniques and Application of Sentiment Analysis", Communication of ACM, vol. 56.No.4, April 2013.
- [15] Ahmad Ashari, Iman Paryudi, A Min Tjoa, "Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 11, 2013.
- [16] Ajayi Adebawale, Idowu S.A, Anyaehie Amarachi A., "Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-3, July 2013.
- [17] Wikipedia, "k-Nearest Neighbor Algorithm," Available at: [http://en.wikipedia.org/wiki/K\\_nearest\\_neighbor\\_algorithm](http://en.wikipedia.org/wiki/K_nearest_neighbor_algorithm).
- [18] V. Garcia, C. Debreuve, "Fast k Nearest Neighbor Search using GPU", IEEE, 2008.
- [19] Bo Pang and Lillian Lee, Shivakumar Vaithyanathan "Thumbs up? Sentiment Classification using Machine Learning Techniques", Proceedings of EMNLP 2002, pp. 79-86, 2002.
- [20] Abdullah Dar\*, Anurag Jain, "Survey paper on Sentiment Analysis: In General Terms", International Journal of Emerging Research in Management & Technology ISSN: 2278-9359 (Volume-3, Issue-11).