

# Review on Developing Corpora for Sentiment Analysis Using Plutchik's Wheel of Emotions with Fuzzy Logic

Dhanashri Chafale<sup>1\*</sup> and Amit Pimpalkar<sup>2</sup>

<sup>1\*,2</sup>Department of Computer science & Engineering, G.H.R.A.E.T, Nagpur University, India.

[www.ijcaonline.org](http://www.ijcaonline.org)

Received: 13 Oct 2014

Revised: 20 Oct 2014

Accepted: 28 Oct 2014

Published: 31 Oct 2014

**Abstract** - Internet is the global system which is increasing day by day with a faster rate. With the increasing internet, social networking increases and people started to share information through different kinds of social media. In recent years several efforts were devoted for mining opinions and sentiments automatically from natural language in social media messages, news and commercial product reviews. This task involves a deep understanding of the explicit and implicit information, conveyed by the language. Most of these approaches refer to annotated corpora. The use of Opinion mining is to identify and extract the information, which is in the subjective form from the internet. This can be done with the help of data, required for processing. The methods used are natural language processing, text analysis. Sentiments are also extracted from the feedbacks. Feedback is important for selling or purchasing any product. While shopping whenever someone wants to choose any product, the opinion of others will always help him/her to choose the best product. But it is very difficult for customer to read thousands of reviews at a time and it also creates confusion. So some data mining techniques must be applied to solve these problems. Sentiment analysis also helps in identifying the attitude of the person. In our work, we present a system which develops a corpus for opinion and sentiment analysis. We will take the product reviews and classify them as positive, negative and objective. The system will further classify the positive and negative sentiments into emotions using Plutchik's wheel of emotions and makes a dictionary. It uses fuzzy logic approach for prediction and generates output.

**Keywords:** Sentiments, Sentiment Classification, Opinion Mining, Corpora for Sentiment Analysis

## I. INTRODUCTION

Mining opinions and sentiments from natural language is a very difficult task. It involves a very intense understanding of most of the explicit and implicit information which is conveyed by language structures. The availability of a dynamic corpus which contains the data generated by user, such as reviews for products or polling data, makes necessary to deal with the affective information conveyed by expressive texts reflecting spontaneous user responses. The large and growing amount of available information in the Social Web fosters the proliferation of business and research activities around the relatively new fields of opinion mining and sentiment analysis. Big Data is trending research area in computer Science which contains sentiment analysis as one of its most important part. Big data is the large amount of data which is available easily on web, remote sensing data, Social media, etc. in form of structured data, semi-structured or unstructured data. The unstructured data may have the forms like news articles, tweets, reviews for movies and products etc. We can use this large data for sentiment analysis.

Sentiment analysis is the opinion mining which is used in identifying the text on the web. It is nothing but to get the real voice of people for specific product, services, movies, news, issues etc. The aim of sentiment analysis is to determine the attitude of a person with respect to some topic. Sentiment analysis can be done at three different levels as, sentence level, document level and entity or attribute level. The attitude may be a person's judgment

for the particular product. Feedback or opinion is very important for customer as well as producer because most of the people purchase or sale the products online. Individual consumers may want the opinions of existing users for the product before

purchasing it. Numbers of web sites are available which gives information about product reviews. But, for customers it is very difficult to read these huge numbers of comments at a time. This creates confusion in mind, due to which lot of time is required to take decision. So, the only option is Data Mining which is used to mine opinion and perform sentimental analysis on this large data.

The main task of this system is gathering the more number of reviews available on different online websites. There are many websites available that are used for online products selling, like amazon, flipkart etc. Gathering overall detail of that particular product, the polarity of the given text is checked at the document level sentences. The result gives confirmation, whether the contents of the documents are positive, negative or neutral. Then it uses Plutchik's wheel of emotions to further categories sentiments into eight basic emotions: i.e. joy, trust, disgust, surprise, anticipation, anger, fear and sadness. This wheel of emotions is invented by Robert Plutchik. The basic emotions can be divided into two polarities i.e. joy opposites to sadness, anger opposites to fear, trust opposites to disgust, and surprise opposites to anticipation. Then each emotion can be further divided into three degrees, for example, serenity is a lesser degree of joy and ecstasy is a more intense degree of joy. The eight

Corresponding Author: *Dhanashri Chafale*

basic emotions can combine to form a new emotion. For example, joy and trust can be combined to form love emotion. However, joy, trust, and anger are combined to form a new emotion called jealousy.

The main aim of the system is analysis of the sentiments for the reviews given on product available on online shopping websites. The input data is collected as reviews for products from the online shopping websites and the social networking sites, because the comments for products are posted on them. It also leads to comparison between any two products and identifies the best product between the two.

## II. LITERATURE REVIEW

Sentiment Analysis is basically used to express sentiment of the individual person. Current state of the art in sentiment analysis classified sentiments into two categories positive, negative. Some works classified sentiments as positive, negative and also in a third category objective (or neutral).

Cristina Bosco, etal [1] works on development of a corpus for opinion and sentiment analysis and presented as a case study Senti- TUT, an ongoing project for Italian aimed at investigating sentiment and irony about politics in social media. They developed the two corpora for twitter, namely TWNEWS and TWSPINO using political tweets. TWNEWS corpus had been extracted by applying filters based on time and metadata, aimed at selecting posts where a variety of opinions about politics is represented. TWSPINO is composed of 1,159 messages from the Twitter section of Spinoza a very popular Italian blog of posts with sharp satire on politics. They extracted posts published from July 2009 to February 2012 and removed advertising (1.5%). Amit Pimpalkar, etal [2] developed a system which shows the comments and reviews for products. They determined the polarity of sentiments of the person from the review and comments. They also identify different smiley occurred in the comments and reviews. Then the comparison between two different products was done with the help of reviews which were identified from the online resources. This comparison leads to find the best product. They used Sentiwordnet and smiley's dictionary for determining the scores of words present in the comment. Classification of Sentiments of words was done in three categorize as, positive, negative and objective. The rule based and fuzzy logic approach was used to give the output.

Rathawut Lertsuksakda, etal [3] developed a model namely Hourglass of emotions to tag Thai language stories using Plutchik's wheel of emotions. They reviewed the adopted computational representation of emotions the so-called Hourglass of Emotion. They also proposed a construction of Thai sentiment resource based on such representation for Thai sentiment term tagging. The hourglass of emotions improved upon Plutchik wheel of emotions where Plutchik introduced eight basic emotions: i.e. joy, trust, disgust, surprise, anticipation, anger, fear and

sadness. There are also three degrees of each emotion. Aditi Gupta, etal [4] developed a system in which is useful information is collected from the twitter website and sentiment analysis is performed on the tweets regarding the smart phone war. The developed system uses the efficient scoring system which is used for predicting the user's age. A well trained Naive Bayes Classifier is used to predict user gender. Tweet were labeled with a sentiment using Sentiment Classifier Model which helped in analyzing the data which is based on various consumer parameters such as gender, age group and location. A. Mudinas, etal [5] developed the system in which lexicon and learning based approach combined for concept-level sentiment analysis. Hemalatha, etal [6] developed a system in which pre-processing and machine learning techniques combined to collect tweets from social networking sites. The noise in the data was removed using preprocessing techniques. Machine learning techniques were applied on those tweets which improve business intelligence by providing some prediction for decision making. Classification of results of specific issue analysis was done as Positive, Negative and Neutral. They studied three machine learning algorithms and developed a machine learning tool for sentiment analysis. They also compared the size of file before and after applying the tool. Aurangzeb Khan, etal [7] presented a tool for giving the online review which was a sentence level statement.

Jalaj S. Modha, etal [8] discussed about exiting approaches, methods etc. for performing sentimental analysis on unstructured data available on web. Previously, Sentiment Analysis concentrated for subjective statements or on subjectivity and it just overlooked objective statements which carry sentiment(s). They proposed a new approach which classifies and handles not only subjective but also objective statements for sentimental analysis. They used the four steps for classification in which first documents are categories in opinionated and non-opinionated sentences and then opinionated sentences were taken to further divide them as subjected and objective. After that both subjective and objective were divided as positive, negative and neutral in separate steps. They have evaluated their experimental results by using information Retrieval matrices such as precision, recall, f-measure and accuracy. G. Vinodhini etal [9] presented a survey which covered the methods and techniques in sentiment analysis including challenges appeared in the field. They compared the various techniques for sentiment classification. The techniques compared were machine learning technique using supervised and unsupervised with different algorithms and the feature based sentiment classification etc. They displayed a graphical result showing comparison of these techniques.

LIU Lizhen etal [10] proposed a feature-based vector model and a novel weighting algorithm for sentiment analysis of Chinese product reviews. The model considered modifying the relationships between words and contained rich sentiment strength descriptions, represented by both adverbs of degree and punctuations. Feature vectors were calculated by using dependency parsing. A

novel feature weighting algorithm was proposed for supervised sentiment classification. The experimental results were used to demonstrate the effectiveness of the proposed method compared with a state of the art method using term level weighting algorithms. In [11,12] presented a novel approach which identified feature specific expressions of opinion in product reviews with different features and mixed emotions. They developed a system that extracts potential features from a review and clusters opinion expressions describing each of the features. It finally retrieved the opinion expression which describes the user specified feature. Their developed system showed improved accuracy over the naïve baseline. They also showed that using supervised classification, the system outperforms the naïve baseline by a huge margin.

### III. CHALLENGES IN SENTIMENT ANALYSIS

Sentiment Analysis/Opinion Mining is a recent subtask of Natural Language processing. There are several challenges in Sentiment analysis field. The first is that, an opinion word which is considered as positive in one situation can be considered as negative in some other situations. Ex.: take the word "long", if a customer said that the battery life of laptop was long, it indicates a positive opinion for laptop. But, if the customer said that the start-up time of laptop was long, then it indicates a negative opinion.

Another challenge is that a person doesn't always express opinions in the similar way. Most of the traditional text processing uses the approach which relies on the fact that minute differences between pieces of text don't change their meaning very much. Ex.: "the product was good" is very different from "the product was not good".

People can be contradictory in using their statements. Some reviews will have both positive as well as negative comments. For example: "the movie bombed even though the lead actor rocked it". This is not much difficult for a human to understand, but for a computer it is not easy to parse.

One drawback of the sentiment analysis using combination of lexicon based and learning based approaches at document level is that reviews with a lot of noise are often assigned a neutral score. The reason for this is that the method fails to detect any sentiment.

### IV. PROPOSED SYSTEM

The focus of the system is to provide analysis of sentiments for any product reviews and also to identify the sentiment of the product efficiently to get information of best product among many products. It collects all the customer reviews for different products which contain the facts and opinions. The subjective sentences are classified into three categories as positive, negative and neutral using the machine learning technique based hierarchical clustering. Then the Plutchik's wheel of emotion is used to further classify the positive and negative sentences into different emotions. For this machine learning based neural

network is used. Association rule based approach is used to classify product feedbacks according to sentiments and the corpus is developed in hierarchical form. Finally the fuzzy logic is used to prediction and gives the best product. The steps are,

- Collecting Product reviews from customers which contains opinions.
- Applying the filtering process and remove noise from sentences using spelling correction, stop word removal etc.
- Classifying sentiments as positive, negative and neutral.
- Again classifying positive and negative sentiments into emotions using Plutchik's wheel of emotions.
- Developing a corpus in hierarchical form.

#### A. Plutchik's Wheel of Emotions

Robert Plutchik has invented a wheel of emotions. He suggested 8 primary emotions: joy and sadness; anger and fear; trust and disgust; and surprise and anticipation. These four pairs of opposite emotions show the 8 basic emotions. Additionally, his model shows connectivity between the ideas of an emotion circle with a color wheel. Like that of the case of colors, primary emotions can also be expressed at different degrees of their intensities, for each emotion there are three degrees. For example, serenity is a lesser degree of joy and ecstasy is a more intense degree of joy. Plutchik emotions can be mixed with one another to form a new emotion. For example, joy and trust are combined to form emotion, love. Likewise, joy, trust, and anger are combined to be jealousy.

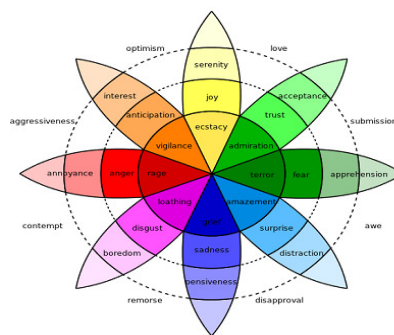


Fig. 1. Plutchik's wheel of emotions.

#### B. Phases of the System

- 1) **Data collection:** The process of gathering the information from different resources is called as data collection. The proposed system will collect the data from the reviews for products given by the customers. We will collect them from the online shopping or product review websites.
- 2) **Pre-processing:** The pre-processing is the process of removing noise which is an unwanted data from the customer's feedbacks. The unwanted data may be the articles, stop words, etc. Short text classifier technique is then used for classification and representation, because it deals with the short text

messages and the feedbacks are not sentences they are short messages.

3) **Clustering:** To categories the feedbacks into positive, negative and neutral clustering technique called hierarchical clustering will be used. Hierarchical clustering is used to build hierarchy of clusters to make the corpus according to plutchik's wheel. Clustering starts with the positive, negative and neutral tags, where tags refer to the particular group which contains sentiments of same type. For ex. Positive tag contains all +ve sentiments. For further processing only positive and negative sentiments will be considered.

4) **Sentiments classification:** The positive and negative feedbacks will be further classified into sentiments. We are going to use plutchik's wheel of emotions for the purpose. Using this classification the corpus will be developed. For this classification machine learning based neural network will be used. Here we are going to use artificial neural network which is a computational model capable of machine learning. We will have to process multiple inputs as there are both positive and negative tags and want to produce more than a single output. Hence we are using neural network which is accept multiple inputs and give multiple outputs. It is used to classify positive and negative sentiments into emotions.

5) **Prediction:**

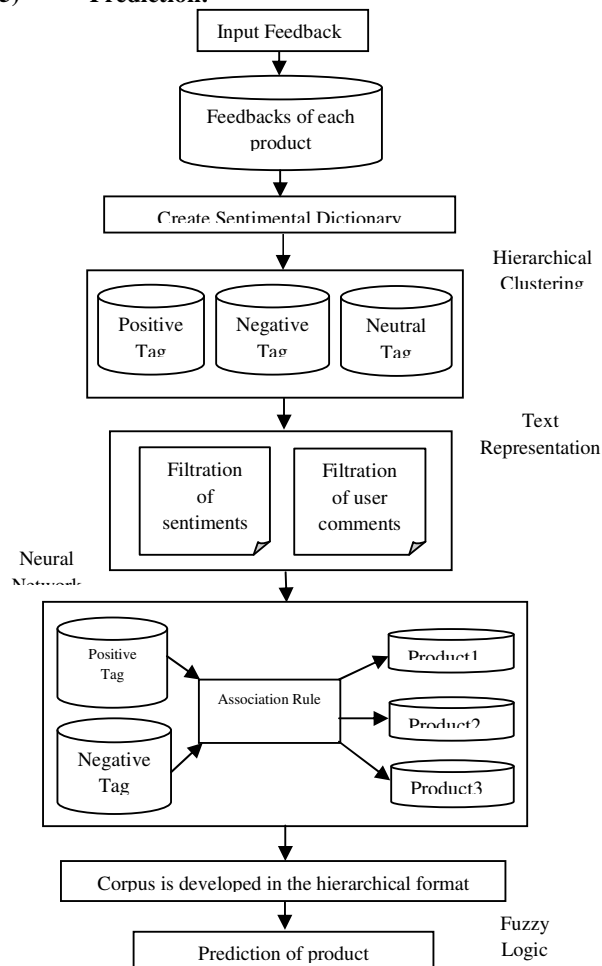


Fig. 2. Work flow of system

In prediction phase, the fuzzy logic will be used to predict the product. Fuzzy logic is basically many valued logic. It deals with the approximate values instead of fixed or exact. So making use of some fuzzy approximations we will be able to compare products and will make the decision whether the product is good or bad. It will give the output as which product is good and which is bad according to customer's opinion.

The Figure below shows the work flow of the proposed system. It gives the brief idea about how the proposed system will work.

## V. CONCLUSION:

There are wide varieties of applications of sentiment analysis in various systems which include classifying and summarizing reviews and other real time applications. The different social media like facebook, twitter, online review websites like product review, movie reviews etc helps user to find answers about anything they are interested in. The sentiment analysis for the products review will help the customer to choose the best product. Also it will help the developer or company to remove the disadvantages of their product or services and re-design them according to customer's need.

## ACKNOWLEDGMENT

The authors would like to thank fellows of IJCSE for their reviews on this paper. I am grateful to my guide Prof. Amit Pimpalkar for his valuable suggestions and encouragement. Special thanks to the authors of the reference papers which helps me to understand the different techniques.

## REFERENCES

- [1] Cristina Bosco, Viviana Patti and Andrea Bolioli, "Developing corpora for sentiment analysis and opinion mining: the case of irony and Senti-TUT", IEEE Intelligent Systems, **2013**.
- [2] Amit Pimpalkar, Tejashree Wandhe, M. Swati Rao and Minal Kene "Review of Online Product using Rule Based and Fuzzy Logic with Smiley's", International Journal of computing and technology (IJCAT), Volume 1, Issue 1, February **2014**.
- [3] Rathawut Lertsuksakda, Ponrudee Netisopakul and Kitsuchart Pasupa "Thai Sentiment Terms Construction using the Hourglass of Emotions", 6th International Conference on Knowledge and Smart Technology (KST), **2014**.
- [4] Aditi Gupta, Karthik Sondhi, Nishit Shivhre and Raunak Kumar, "Sentiment Analysis for Social Media", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July **2013**.
- [5] A. Mudinas, D. Zhang, and M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis", Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, no.5, pp. 1-8, **2012**.
- [6] Hemalatha, Dr. G. P Saradhi Varma and Dr. A. Govardhan, "Sentiment Analysis Tool using Machine Learning Algorithms", International Journal of Emerging Trends &



- Technology in Computer Science (IJETTCS), Volume 2, Issue 2, March – April 2013.
- [7] Aurangzeb Khan, Baharum Baharudin and Khairullah Khan, "Sentiment Classification Using Sentence-level Lexical Based Semantic Orientation of Online Reviews", Trends in Applied Sciences Research, Vol. 6, pp. 1141-1157, July 2011.
- [8] Jalaj S. Modha, Gayatri S. Pandi, and Sandip J. Modha, "Automatic Sentiment Analysis for Unstructured Data", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 12, December 2013.
- [9] G. Vinodhini and RM. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012.
- [10] Liu Lizhen, Song Wei, Wang Hanshi, Li Chuchu and Lu Jingli, "A Novel Feature-based Method for Sentiment Analysis of Chinese Product Reviews", in proceedings of ICT Management, China Communication, pp. 154-164, March 2014.
- [11] Subhabrata Mukherjee and Pushpak Bhattacharyya, "Feature Specific Sentiment Analysis for Product Reviews", Dept. of Computer Science and Engineering, IIT Bombay, 2011.
- [12] Jyoti Bala and Renu Dhir, "A Novel Hybrid Technique for Sub-pixel Edge Detection using Fuzzy Logic and Zernike Moment", International Journal of Computer Sciences and Engineering, Volume-02, Issue-04, Page No (26-31), Apr - 2014, E-ISSN: 2347-2693

#### AUTHORS PROFILE

---



Dhanashri D. Chafale, pursuing M.Tech in Computer Science and Engineering branch from G.H. Rasoni Academy of Engineering & Technology, Nagpur, India. The research interest lies in Data mining.



Prof. Amit Pimpalkar, Assistant Professor at G.H. Rasoni Academy of Engineering & Technology, Nagpur, India. The research interest lies in Data mining, Image processing, Natural Language Processing and Artificial Intelligence.