
Sentiment Analysis on Product Reviews Using Plutchik's Wheel of Emotions with Fuzzy Logic

Dhanashri Chafale*

Department of Computer Science & Engineering
G.H.R.A.E.T, Nagpur University, India

Amit Pimpalkar

Department of Computer Science & Engineering,
G.H.R.A.E.T, Nagpur University, India.

Abstract

Internet is the day by day increasing global system. In recent years number of efforts were devoted for mining opinions and sentiments automatically from natural language in social media messages, commercial product reviews, news and movie reviews. This task includes understanding explicit and implicit information conveyed by the language deeply. Most of these approaches refer to annotated corpora. The opinion mining is used for identification and extraction the information, which is in the subjective form and collected from the internet. This can be done by making use of data required for processing. The methods such as natural language processing, text analysis etc are used. Sentiments can also be extracted from the feedbacks. Feedback is important for purchasing or selling any product. While shopping whenever one wants to choose a product, the opinion of others may be useful to him/her to choose the best product. But in case of customers, usually it is a difficult task to read thousands of reviews at a time and also it is likely to be a time consuming process. It also creates confusion. Therefore as a solution for this, some data mining techniques must be applied. One more advantage of sentiment analysis is it 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 collect the product reviews from any of the available product review website and classify them as positive, negative and neutral sentiments. The system will further classify the positive and negative sentiments into emotions for which it uses the concept of Plutchik's wheel of emotions and makes a dictionary. It uses fuzzy logic approach for prediction and generates output.

Keywords: *Opinion Mining, Sentiment Classification, Sentiments, Corpus for Sentiment Analysis.*

***Author for correspondence** ghanashrichafale@gmail.com

1. Introduction

The task of mining sentiments and opinions from natural language is difficult one. It involves an intense understanding of most of the implicit and explicit information which is conveyed by structure of language. The availability of a dynamic corpus contains the user generated data, such as reviews for products or polling data. The large and growing amount of information which is

available in the Social Web fosters the proliferation of business and research activities around the relatively new fields of sentiment analysis and opinion mining. Big data is the large amount of easily available data on web, Social media, remote sensing data, etc. in form of structured data, semi-structured or unstructured data. We can use this large data for sentiment analysis.

Sentiment analysis is the opinion mining used on the web for identifying the text. It is nothing but to get the real voice of people for specific product, services, movies, news, issues etc. Sentiment analysis can be done at various levels as, sentence level, document level and entity or attribute level. The attitude of a particular person may be his/her judgment for the particular product. Opinion or Feedback is very important for consumer as well as producer because most of the people sale or purchase the products online. Individual consumers may want the opinions of already existing users for the product before purchasing it. There are many web sites giving information about product reviews. But, for customers it is somewhat a difficult task to read these huge numbers of comments at a time. This creates confusion in mind, due to which comparatively more time is required to take decision. So, the only option is using Data Mining which mines opinion and performs sentimental analysis on this large data.

The main task of this system is gathering the reviews in large number that are available on different online websites. The websites are available for online products selling, like Amazon, flipkart etc. Gathering overall detail for the particular product, the polarity of the given text is checked at the document level sentences. The result gives confirmation, about the contents of the documents whether it is positive, negative or neutral. Then it uses a technique called “Plutchik’s Wheel of Emotions” to categories sentiments further into eight basic emotions: i.e. joy, disgust, trust, surprise, anticipation, fear, anger and sadness. This wheel of emotions is first invented by the scientist 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 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 jealousy.

The aim of the system is analysis of the sentiments for the online available reviews given on product from online shopping websites. The input data is collected as reviews from the online shopping websites because the comments for products are posted there. It compares between products and identifies the best product.

2. Related Work

Basically, Sentiment Analysis is used to express individual person’s sentiment. According to current state of the art sentiment analysis is used to classify sentiments into two categories positive and negative. Some works classified them into as positive, negative and also in one more category as objective (or neutral).

Pimpalkar et al. [1] developed a system that shows the comments and feedbacks/reviews for products. They determined the polarity of sentiments for the comments of the person. After this the comparison between two different products was done using comments 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 categories as, positive, negative and objective. The rule based and fuzzy logic approach was used to give the output. Lertsuksakda et al. [2] developed a model, Hourglass of emotions to tag Thai 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's wheel of emotions where Plutchik introduced eight basic emotions: i.e. joy, disgust, surprise, anticipation, trust, anger, fear and sadness. There are also three degrees of each emotion. Lizhen et al. [3] 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. Bosco et al. [4] 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%). Gupta et al. [5] 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. Hemalatha et al. [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. Modha et al. [7] 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 categorized in opinionated and non-opinionated sentences and then opinionated sentences were taken to further divide them as subjective 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.

Mudinas et al. [8] developed the system in which lexicon and learning based approach combined for concept-level sentiment analysis. Vinodhini et al. [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. Mukherjee et al. [10] 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. Nitin et al. [11] studied the problem of identifying comparative sentences in text documents. A supervised learning approach is used for identifying comparative sentences from text documents.

3. Challenges in Sentiment Analysis

Sentiment Analysis or Opinion Mining is the recent subtask of NLP (Natural Language processing). Sentiment analysis field are having few challenges. The first is that, a word containing opinion which is considered positive in one situation is not necessarily positive everywhere as it can be considered negative in some other situations. Ex.: take the word “long”, suppose a customer says “The battery life of laptop is long”. This statement indicates a positive opinion for that laptop. But, in some other situation, the customer said that, start-up time of laptop was long, and then this is a negative opinion.

Second thing is that none can always express opinions in the similar way. Almost all of the traditional text processing makes use of the approach that relies on the fact that minute differences between pieces of text not necessarily change their meaning very much. Ex.: “The product is good” is so much different from “The product is not good”.

Sometimes, people may use contradictory 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 so difficult for a human being to understand it. But for a computer/machine it is not so easy to parse. A well known drawback of the sentiment analysis using combination of lexicon based and learning based approaches at document level is, the 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.

4. Proposed System

The focus of the system is on providing analysis of sentiments for product reviews and also identifying the sentiment of the product efficiently for getting information of best product among many products. First thing is it collects all the customer reviews for different products that contain the facts and opinions. The basic classification subjective sentences is done into three

categories as positive, negative and neutral by making use of the Hierarchical clustering. Then, the Plutchik's wheel of emotion is used. It further classifies the positive and negative sentences into different Plutchik's emotions. For this work machine learning based Neural network technique is used. Along with this the corpus showing feedbacks classification is developed in hierarchical form. The corpus will contain detailed classification. Finally the fuzzy logic is used for prediction purpose and gives the best product.

Objectives of the proposed system are,

- a) Collection of product reviews from customers/users which contains opinions.
- b) Applying the filtration process and remove unwanted data (noise) from sentences using stemming, stop word removal etc.
- c) Classifying sentiments as positive, negative and neutral.
- d) Again classifying positive and negative sentiments into various emotions using Plutchik's wheel of emotions.
- e) Developing corpus in hierarchical format.

A. Plutchik's Wheel of Emotions

Robert Plutchik invented a wheel of emotions. He suggested eight primary emotions: joy opposite to sadness. Similarly, anger opposite to fear; trust opposite to disgust and surprise opposite to anticipation. These four opposite emotion pairs, show the 8 basic emotions. Additionally, Plutchik's model shows connections between the ideas of circle of emotions using a color wheel. Like 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 less intense degree of joy and ecstasy is a more intense degree of joy. Plutchik's emotions can be mixed with one another forming a new different emotion. For example, combination of joy and trust resulted to form a new emotion 'love'. Likewise, joy, anger and trust are combined and form jealousy.

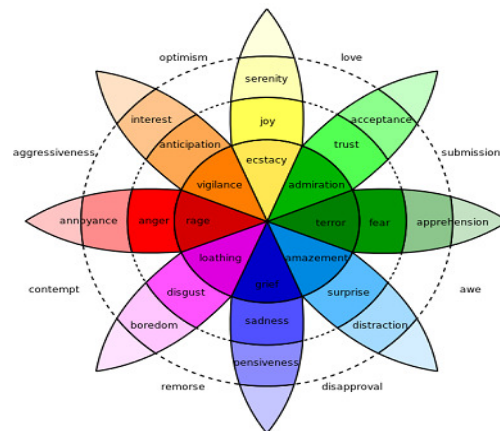


Fig. 1: Plutchik's Wheel of Emotions

Phases of the System

- 1) Data collection: The process of collecting information data collection. Our proposed system is going to collect the data from the comments given for products by the customers. We use online shopping websites like amazon, flipkart and many more available there for collecting user's comments. We can take different types of product and even multiple companies for the same product and collect reviews for each of them.
- 2) Pre-processing: The pre-processing means the process which removes noise that is an unwanted data from the customer's comments for product. The noise may be in the form of articles, stop words, etc. Different procedures are used for stemming; by making use of them the comments will be filtered. Short text classifier is used for classification and

representation; because this technique particularly deals with the short text messages and the feedbacks/comments are short text messages.

- 3) Clustering: To show the categorization of feedbacks of the users as positive, negative and neutral, one of the clustering techniques will be used. Here for our project we are going to use a clustering technique called hierarchical clustering. Hierarchical clustering is used to build hierarchy of clusters and using this concept the corpus is developed according to Plutchik's wheel. Clustering starts with the positive tags, negative tags and neutral tags, where tags refer to the particular group which contains sentiments of same type. For ex. negative tag contains all -ve sentiments. For further processing we will consider only negative and positive sentiments.
- 4) Sentiments classification: In this phase the positive and negative feedbacks will be classified further into different emotions. We will make use of Plutchik's wheel of emotions for this classification. The corpus will be developed on the basis of this classification. For this classification Neural network which is a machine learning based technique, will be used. Here we have decided to use Artificial Neural Network which is a computational model having capability of machine learning. We will have to process multiple inputs because there are positive as well as negative tags and also we want to produce more than one output. For this reason we are going to use neural network which accepts multiple inputs and produce multiple outputs. It will classify positive and negative sentiments into emotions.
- 5) Prediction: Here in the last phase called prediction, the fuzzy logic will be used for the prediction purpose that is to predict better product among various. Fuzzy logic basically, is a many valued logic. Fuzzy logic deals with the approximate values instead of exact or fixed. Therefore by making use of some of the fuzzy approximations we will be able to compare products of different companies. From this process we will make the decision as which product is better or best. It will give the output as if the given product is good or bad. And also it will predict the best product which is our main aim.

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

B. Implementation Details

The product reviews are collected from the online product review websites. For this we have taken the dataset containing user's feedback for various products. We requested for this data from the Amazon, the online shopping website and stored in a database. We maintained a sentiment words dictionary which contains positive words dictionary and negative words dictionary. All the comments for the selected products are shown first and then we have to choose a comment from the list to further process it. When any one comment is selected it starts processing on that comment.

First of all the stemming and stopword techniques are applied on it. This filtering removes almost all unwanted noise from comment. The filtered comment is then split to get the separate words for comparing. Then each single word is compared with the sentiment words dictionary. If the word is matched with the positive or negative dictionary then it is placed in the

corresponding box, that is positive word in positive words text and in the same way negative words are placed. The comparison is done between number of positive word and number of negative words in a given comment. The condition is checked whether the positive words are more or negative and accordingly the comment is decided to be positive or negative. If both the positive and negative words are same or if there are no positive or negative, the comment is treated as neutral comment.

The implemented work is described above briefly. The total project consists of five modules. We have developed the two modules and described them here shortly. The further development is in progress. Until now we have done the data collection and preprocessing on the comments. And we are selecting only a single comment at a time and showed whether it is positive, negative or neutral. Now we will work in the same way for all comments for a particular product. In future we will classify the comments of users according to Plutchik's wheel of emotions. And also show the prediction process that is compare the product and give the best product as a result.

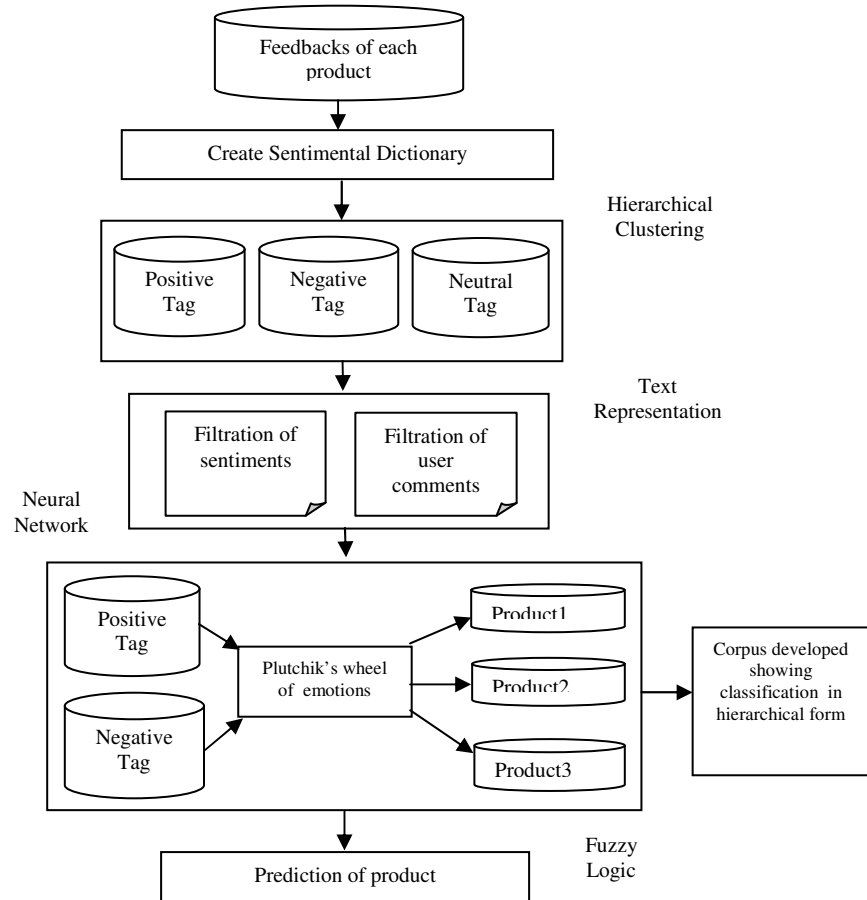


Fig. 2: Workflow of system

5. Conclusion

The wide varieties of sentiment analysis applications are there in various systems that include classifying and summarizing reviews. We used here a hierarchical clustering and showed the hierarchy of sentiments in tree form. 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. The use of Plutchik's wheel of emotions will provide the real emotional view of comments.

References

- [1] Pimpalkar, A., Wandhe, T., Kene, M., & Rao, M. S. (2014). Review of online product using rule based and fuzzy logic with Smiley's. *International Journal of Computing and Technology*. 1(1), 39-44.
- [2] Lertsuksakda, R., Pasupa, K., & Netisopakul, P. (2014). Thai sentiment terms construction using the Hourglass of emotions. *In: Proceedings of 6th International Conference on Knowledge and Smart Technology (KST)*, pp. 46-50).
- [3] Lizhen, L., Wei, S., Chuchu, L., Hanshi, W., & Jingli, L. (2014). A novel feature-based method for sentiment analysis of Chinese product reviews. *In: Proceedings of ICT Management, China Communication (ICTM-2014)*, pp. 154-164).
- [4] Bosco, C., Bolioli, A., & Patti, V. (2013). Developing corpora for sentiment analysis and opinion mining: the case of irony and Senti-TUT. *IEEE Intelligent Systems*.
- [5] Gupta, A., Sondhi, K., Kumar, R., & Shivhre, N. (2013). Sentiment analysis for social media. *International Journal of Advanced Research in Computer Science and Software Engineering*. 3(7), 216-221.
- [6] Hemalatha, G. A., & Saradhi Varma, G. P. (2013). Sentiment analysis tool using machine learning algorithms. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*. 2(2), 105-109.
- [7] Modha, J. S., Modha, S. J., & Pandi, G. S. (2013). Automatic sentiment analysis for unstructured data. 3(12), 91-97.
- [8] Mudinas, A., Levene, M., & Zhang, D. (2012). Combining lexicon and learning based approaches for concept-level sentiment analysis. *In: Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining* (pp. 1-8).
- [9] Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2(6), 281-292.
- [10] Mukherjee, S., & Bhattacharyya, P. (2011). Feature specific sentiment analysis for product reviews. Dept. of Computer Science and Engineering, IIT Bombay.