Sentence Based Sentiment Analysis on Users Reviews

Antara Ayyar[1], Pradnya Awate[2], Sarika Daundkar[3], Kranti Rachure[4]
Information Technology[1, 2, 3, 4], MIT Academy Of Engineering, Alandi Pune[1, 2, 3, 4]
Email: antaraayyar@gmail.com[1], pradnyaawate8@gmail.com[2], sarikadaundkar6@gmail.com[3]
krantirachure@gmail.com[4]

Abstract- Sentiment Analysis (SA) is considered to be a field of research in text mining field. SA can be defined in a word as the treatment of opinions, sentiments using computations. It is also considered as an alternative research technique to collect and analyze textual data on the internet. Systematic and computational evaluation of textual content is done by Sentiment analysis, hence it is also called as a data mining technique. This paper briefs about the basics of SA, existing techniques related to SA. It also covers the approaches, applications and the challenges involved in sentiment analysis. The main goal is to perform sentiment analysis by categorizing users comments as positive or negative and from which conclusions can be made.

Index Terms-Sentiment Analysis, machine learning algorithm.

1. INTRODUCTION

People express their views/opinions freely in today's world and these views/opinions become very helpful source of information for making various policies /decisions. Due to the advent of World Wide Web, social media has become a good source containing opinionated text. This domain of computer science is also called as Sentiment Analysis, Subjectivity Analysis, Sentiment classification and Opinion mining. It is the study of person's views, emotions or opinions associated with some products, some events, some places and their features, etc. It is the field of study that analyzes people's opinions, emotions, evaluations, sentiments, attitudes towards entities such as individuals, events, products, services, organizations and their attributes as well.[1]

However, the term *sentiment analysis and opinion mining are very popular in industrial and research field.* They represent the same area of study basically. Opinion mining and sentiment analysis mainly focuses on opinions which imply positive or negative sentiments. The sentiments that we find within comments, feedback or critiques provide useful indicators for different purposes. They can be categorized either into two categories i.e. positive and negative; or into an n-point scale, e.g., excellent, very good, good, poor, bad, very bad, satisfactory. In short, a sentiment analysis task can be interpreted as a classification of task where each category or classification represents a sentiment.

Sentimental Analysis is entirely related to extracting and analyzing the views of people towards various products. An example of this could be Tweets, Facebook comments, blogs, product reviews, and so on. Sentiment Analysis can be used in branches of computer science like Natural Language Processing, Text Mining, Machine Learning, Information Theory and Coding. Using various approaches, methods, techniques and models of defined branches, we can categorize our data which is unstructured (in form of News, Comments, Tweets, Articles, Blogs and reviews on products, movies etc.) into positive, negative or neutral sentiment according to the sentiment is expressed in them. For example, in marketing sector it helps in judging the success of new deployed product in the market, it is useful to determine which versions of a product or service are popular and which features of the product are liked or disliked by users.

2. EXISTING TECHNIQUES AND APPROCHES

2.1. *Machine learning techniques*

Machine learning techniques are the most used techniques for the sentiment analysis for categorizing sentences or documents into positive, negative or neutral categories.
The Machine learning techniques are mainly classified into two basic types:-

2.1.1. *Supervised Machine Learning Techniques*

Sentiment classification is usually categorized as a two-class classification problem i.e. positive and negative. Normally product reviews are used as training and testing data. Online reviews do have reviewers that assign rating scores, e.g., 1-5 stars is used as a scale for giving the ratings. For instance, a review with 4 or 5 stars falls in the category of a positive review, and a review with 1 to 2 stars falls in the category of a negative review. 3 stars is usually for the neutral class. Sentiment classification is essentially known as the text classification problem. Traditionally text

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

classification basically classifies documents of various topics, e.g., politics, sports and sciences. In such a classification, topic-related words are the key features. In sentiment classification, opinion or sentiment words that indicate positive or negative opinions are rather more important, e.g., excellent, amazing, great, bad, horrible, worst, etc. Types of supervised learning method used can be "naïve Bayes classification" or "support vector machines (SVM)".

A Naive Bayes classifier is said to be a probabilistic classifier based approach. Whereas, Support vector machines is an approach based on the theory of statistical learning.

2.1.2. Unsupervised Machine Learning Techniques

This is another approach in Machine Learning. Since sentiment words are often the main factor for classification, sentiment words and sentences may also be used for unsupervised manner of sentiment classification. One of the approaches based on unsupervised machine learning method is: classification that is performed based on some fixed syntactic patterns that are most likely to be used to express opinions. The syntactic patterns depends upon part-of-speech (POS) tags.

Another unsupervised approach is the lexicon-based method, which uses a dictionary of words to compute a sentiment score for each document. It takes into consideration associated orientations and strength related to the words. Unsupervised machine learning techniques do not make use training data set for classification. Clustering algorithms like Hierarchical clustering, K-means clustering used to classify data into categories. Semantic Orientation also provides to generate accurate result for classification. [1]

3. LITERATURE SURVEY

Sentiment analysis is a term used for studying users opinion, views or emotions on a particular product, some event, their related features and so on. As mentioned in the Sentiment Analysis and Opinion Mining-Bing Liu .Opinions and its related concepts such as sentiments, evaluations, attitudes, and emotions are the main domains of study of sentiment analysis and opinion mining. This book gives the detail about the concept of sentiment analysis as a whole. It mainly covers the problem of sentiment analysis, the different approaches involved in sentiment analysis from machine learning techniques to sentiment lexicon generation. Not only this it also provides with a comprehensive survey of all important aspects and the latest developments in this field [1].

Many researchers have contributed to a significant amount of research in this field. This section highlights some of the major work done:-

Prof. Ronen Feldman, had presented a tutorial on sentiment analysis, based on the slides from Bing Lu[9]. It briefs about the basics as well as the applications of sentiment analysis. It focuses on the dictionary based and corpus based approach.

Walaa Medhat, Ahmed Hassan, Hoda Korashy have successfully done a detailed study on various approaches to implement sentiment analysis in "Sentiment analysis algorithms and applications: A survey"[3].Their work gives the detailed explanation of the different sentiment classification techniques such as:-machine learning (Support Vector Machine, Naive Bayes, Neural Network, decision tree classifiers, rule based classifiers etc.),lexicon based (corpus and dictionary based).

Jasleen Kaur, Dr. Jatinderkumar R. Saini,proposed a paper -" On Classifying Sentiments and Mining Opinions."[4].This paper tells us about the different sources of data that can be used as an input for analyzing the sentiment. It also mentions about the opinion mining done at document level, sentence level and feature level. It also gives information about various approaches of classifying sentiments.

G.Vinodhini,RM.Chandrasekaran have mentioned many different data sources, approaches, application and tools for sentiment classification in their survey named-Sentiment Analysis and Opinion Mining: A Survey[6].This paper discusses about different Machine learning techniques, concept of semantic orientation, role of negation and it significance while determining the polarity.

4. PROPOSED METHODOLOGY

4.1. *Data collection*

Sentiment analysis makes the use of huge user generated content over the internet. The data source points to queries of user discussions on public forums like private logs, blogs, as well as on product reviews boards through social network sites like Facebook and Twitter. The data log is usually bulky, disintegrated, and disorganized on multiple portals. Feelings and opinions are expressed in various ways including the type of vocabulary used, amount of detailed data given, slangs and lingua variations, context of writing are just a few examples. This makes analysis tedious and almost impossible manually. As the extraction of data is done, it will then be prepared for further

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

analysis. The polarity or say the orientation of an opinion can either be positive, negative or neutral. An opinion can be of a person or organization. If we consider the case of blogs and product reviews, the authors of the posts are usually opinion holders. Opinions of users play a vital role for the improvement of the quality of services rendered as well deliverable enhancement. Review sites, blogs and micro blogs provide a better understanding of the acceptance level of the services and products.

4.2. *Text preparation*

Text preparation is an act of cleaning the data that is extracted before performing analyses. Text preparation usually involves identifying and eliminating the non textual content from the textual dataset, and any data that can tell about the reviewers identity such as: reviewer name, reviewer date, review location. In addition to this, any other content that is irrelevant to the area of study is also discarded from the textual dataset like stop words or words that are irrelevant to the course of analysis.

4.3 Sentiment detection

The third stage in sentiment analysis is sentiment detection. Each sentence is evaluated for its subjectivity. There is a database consisting of words with their positive and negative scores. From the second stage the filtered content is analyzed and the scores are determined for every particular word, by comparing the word in the content with the available database used as training dataset.
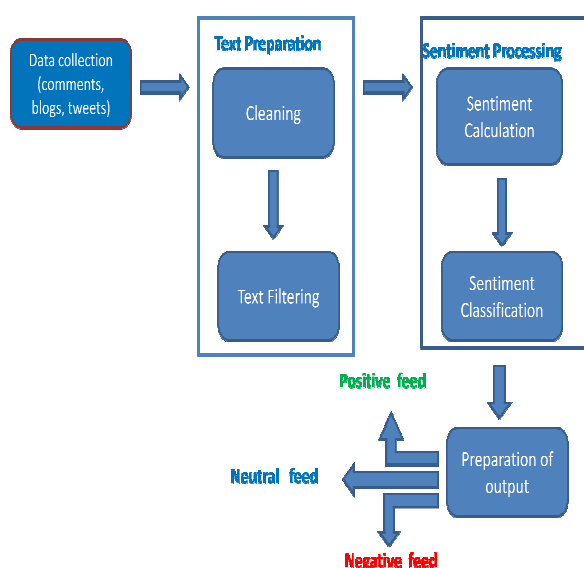

Fig.1. Proposed System Architecture

4.4. *Sentiment classification*

The fourth stage is sentiment classification or polarity classification which classifies each subjective sentence that is present in the textual dataset into groups called classification groups. These groups are mainly represented on two extreme points:-positive, negative.

4.5. *Presentation of output*

The general purpose of the analysis is to convert unstructured fragmented text into meaningful information. Once the analysis is completed, a number of conventional options are used to display the result of text analysis. Chief among them is the use of graphical displays such as pie charts, bar charts and line graphs. The polarity is segmented on colour, frequencies, percentages and size. The format of presentation depends on the research interest.[10]

5. SYSTEM IMPLEMENTATION AND RESULTS

Sentiment analysis is mainly considered as a task of Sentiment Classification. The First step in Sentiment Classification problem is to collect the data and perform the required cleaning and filtering operations on the data. Filtering process is a process wherein the words in comments or tweets is checked for its correct form. The system consists of three databases:-

- Database consisting of dictionary words.
- Database consisting of list of slang words.
- A sentiword database consisting of the individual scores of words.

Firstly, every sentence would be broken down into words. Every word will be an input to the dictionary database. If the word is present there, it would go to the sentiword database directly. If not it would check for its existence in the slang word database. If present in the slang word database, it would then be directed to the sentiwords database for its score. The average is now calculated for every word and the sentiment is detected based on the score. If the positive average score is greater than the negative average score, the sentiment is detected as positive and vice versa.

Calculation for classification:-

- Positive sentiment=Avg(positive scores) > Avg (negative score)
- Negative sentiment=Avg(positive scores) < Avg (negative score)

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

This way sentences would be classified according to the sentiment value it holds. The main aim is to achieve maximum efficiency using this approach.
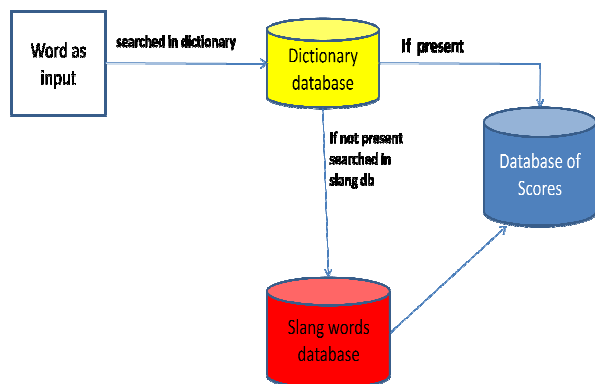

Fig.2. System Work Flow

The system implemented has been tested for various set of inputs. More efficiency is to be achieved in the existing system.
Some of the examples are shown in the below table.

Table 1. Result Table

Sr. no	Sentence	Filtered Word	Result
1	I am a good girl	Good=0.256288065	Positive
2.	It was a bad move	Bad=-0.157995633	negative
3.	we were at last able to buy a car	Able= 0.19086184	positive
4.	Things are right now.	Right= 0.10368217	positive
5.	He was the worst dressed person present	Worst= -0.5568181	negative

6. APPLICATIONS AND USES OF SENTIMENTAL ANALYSIS

(1) You could use SA to: search the web for opinions and reviews of competing laptops. Blogs, opinions, amazon, tweets, etc.
(2) Cross domain applications:- Insights and applications from SA have been useful in other areas

- Politics/political science:

  (1) Evaluation of public/voters' opinions
  (2) Views/discussions of policy

- Sociology:-Analyzing sentiment reactions on blogs can give insight to this process.[5]

7. CONCLUSION

Sentiment Analysis (SA) is an ongoing Field of research in text mining field. SA is the computational treatment of opinions, sentiments and subjectivity of text. The data used in SA are mostly on Product Reviews posted on social networking sites in form of tweets, blogs, comments etc. It is also found that different types of features and classification algorithms are combined in an efficient way. The supervised approach works well for sentiment classification analysis The proposed methodological approach to sentimental analysis, may be adapted for solving of sophisticated sentiment analysis problems. This sentence based approach extracts, filters, analyzes and categorizes the sentences based on its polarity and shows some significant results on a set of few sentences, which can be definitely improved if a words for filtering are directly match with the google's database or by using a dataset of larger size.

8. FUTURE WORK

In future, more work is needed on further improving the performance measures. Sentiment analysis can be applied for new applications. Although the techniques and algorithms used for sentiment analysis are advancing fast, however, a lot of problems in this field of study remain unsolved. The main challenging aspects exist in use of other languages, dealing with negation expressions; produce a summary of opinions based on product features/attributes, complexity of sentence/document , handling of implicit product features. Implementing a system supporting multilingual platform is also one of the major challenge in future.

9. ACKNOWLEDGEMENT

International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

10. REFRENCES

[1]Sentiment Analysis and Opinion Mining, April 22, 2012 Bing Liu, liub@cs.uic.edu

[2] Automatic Sentiment Analysis for Unstructured Data,Jalaj S.Modha,Prof Head Gayatri S. Pandi,Sandip J. Modha, www.ijarcsse.com

[3] Sentiment analysis algorithms and applications: A survey,Walaa Medhat,Ahmed Hassan,Hoda Korashy, www.sciencedirect.com

[4] On Classifying Sentiments and Mining Opinions,Jasleen Kaur et al.,International Journal of Emerging Technologies in Computational and Applied Sciences, 9(3), June-August, 2014, pp., www.iasir.net

[5] Introduction to Sentiment Analysis(ppt).

[6] Sentiment Analysis and Opinion Mining:A Survey;G.Vinodhini,RM.Chandrasekaran www.ijarcsse.com.

[7] Twitter as a Corpus for Sentiment Analysis and Opinion Mining, Alexander Pak, Patrick Paroubek. alexpak@limsi.fr

[8] Preprocessing the Informal Text for efficient Sentiment Analysis, I.Hemalatha, Dr. G. P Saradhi Varma, Dr. A.Govardhan.

[9]Sentiment analysis tutorial(ppt.), Prof. Ronen Feldman, Ronen.Feldman@huji.ac.il.

[10] Marketing Research:The Role of Sentiment Analysis, Meena Rambocas, João Gama.