Automatic Sentiment Analysis of User Reviews

Abinaya.R[1], Aishwaryaa.P[2], Baavana.S[3], UG Students,
Thamarai Selvi.N.D.[4], Assistant Professor,
Department of Computer Science and Engineering,
Easwari Engineering College, Chennai, India.
[1]abinaya2409@gmail.com, [2]aishu.partha@gmail.com ,[3]sbaavana@gmail.com, [4]thamarai.nd@gmail.com

.

Abstract — **Data mining is the process of turning raw data into useful information. The main use of data mining is to fetch the required data and extract useful information from the data and to interpret the data. In the existing system, Bag of Words model is used along with Dual sentiment Analysis in order to classify the reviews as positive, negative and neutral. However, the performance of Bag of Words sometimes remains limited due to some fundamental deficiencies in handling the polarity shift problem. The proposed system uses a dictionary based classification for accurately classifying the reviews as positive, negative and neutral. To enhance the accuracy in the classification of neutral reviews, Support Vector Machine algorithm is implemented. Both the product owner and the user can identify the quality of the product based on the sentiment graph that is generated based on the reviews for each of the product video. A comparative study of the sentiment graphs is performed in order to improve the efficiency of visual representation.**

Keywords: Opinion Mining, Sentiment Analysis, Data Mining

I. INTRODUCTION

Internet has gained its access in each and every domain and has become an integral part in all walks of life. In today's world, people give their reviews over the internet. All these reviews are being stored as huge amount of data for each and every minute. These reviews have an equal importance as that of buying the commodity. Analyzing these reviews will enhance both the end users and the account executive. Sentiment analysis has earned its identification and is used in classifying the reviews.

Sentiment analysis, also called opinion mining, is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. However, they are now all under the umbrella of sentiment analysis or opinion mining.

While in industry, the term sentiment analysis is more commonly used, but in academia both sentiment analysis and opinion mining are frequently employed. They basically represent the same field of study. The meaning of opinion itself is still very broad. Sentiment analysis and opinion mining mainly focuses on opinions which express or imply positive or negative sentiments.

To do an analysis, classification plays a key role in opinion mining. A Classification Algorithm is a procedure for selecting a hypothesis from a set of alternatives that best fits a set of observations.

Opinions are central to almost all human activities because they are key influencers of our behaviours. Whenever there is a need to make a decision, others' opinions are required. In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. Individual consumers also want to know the opinions of existing users of a product before purchasing it, and others' opinions about political candidates before making a voting decision in a political election. In the past, when an individual needed opinions, he/she asked friends and family. When an organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Acquiring public and consumer opinions has long been a huge business itself for marketing, public relations, and political campaign companies.

Opinion summarization summarizes opinions of articles by telling sentiment polarities, degree and the correlated events. With opinion summarization, a customer can easily see how the existing customers feel about a product, and the product manufacturer can get the reason why different stands people like it or what they complain about.

A seller's job can be quite complicated or it can be quite easy. The two contradictory terms define the selling experience, based on the fact as how seller interprets the consumer interests. Unless one is a psychic or knows how to get into others mind the actual demand of the customer's and the product can't be collaborated. Having a right product is important and equally important is to present it before the

right customer (one who actually needs it or is interested in it). The product should put on positive feeling of ownership among the individuals. And such feelings are clearly expressed in opinion mining polls.

II. LITERATURE SURVEY

Rui Xia, FengXu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li [10].In recent years, with the growing volume of online reviews available on the Internet, sentiment analysis and opinion mining, as a special text mining task for determining the subjective attitude (i.e., sentiment) expressed by the text, is becoming a hotspot in the field of data mining and natural language processing [2], [7], [8]. Sentiment classification is a basic task in sentiment analysis, with its aim to classify the sentiment (e.g., positive or negative) of a given text. The general practice in sentiment classification follows the techniques in traditional topic-based text classification, where the Bag-of words (BOW) model is typically used for text representation. In the BOW model, a review text is represented by a vector of independent words.

Although the BOW model is very simple and quite efficient in topic-based text classification, it is actually not very suitable for sentiment classification because it disrupts the word order, breaks the syntactic structures, and discards some semantic information. Consequently, a large number of researches in sentiment analysis aimed to enhance BOW by incorporating linguistic knowledge. However, due to the fundamental deficiencies in BOW, most of these efforts showed very slight effects in improving the classification accuracy. One of the most well-known difficulties is the polarity shift problem. Polarity shift is a kind of linguistic phenomenon which can reverse the sentiment polarity of the text. Negation is the most important type of polarity shift.

A simple yet efficient model, called dual sentiment analysis (DSA), is developed to address the polarity shift problem in sentiment classification. By using the property that sentiment classification has two opposite class labels (i.e., positive and negative), we first propose a data expansion technique by creating sentiment reversed reviews. The original and reversed reviews are constructed in a one-to-one correspondence. Thereafter, we propose a dual training (DT) algorithm and a dual prediction (DP) algorithm respectively, to make use of the original and reversed samples in pairs for training a statistical classifier and make predictions.

To reduce DSA's dependency on an external antonym dictionary, we finally develop a corpus-based method for constructing a pseudo-antonym dictionary. The pseudo antonym dictionary is language-independent and domain-

adaptive. It makes the DSA model possible to be applied into a wide range of applications.

III. PROPOSED SYSTEM

The system consist of three major steps

1. Reviews Extraction
   i. Creation of user interface and uploading videos
   ii. Word clustering
   iii. Preprocessing

2. Reviews Evaluation
   i. Selection of hypotheses set
   ii. Classification

3. Sentiment graph visualization.

The first step of the module involves creation of web based interface. Upon creation, the admin will login and upload videos which are in MP4 format. Then, the user will login and view videos. He shall then give reviews for the video which will get stored in the database. The stored reviews are then processed with the help of word clustering and pre processing.

In review evaluation, refining of the text in the form of relevant phrases, words with the help of classification is performed. Classification is based on the supervised learning. The algorithm classifies the words given by the user into positive, negative and neutral. The words that do not have any meaning is considered as a hypothetical words. Such kind of words is processed with the help of an iterative induced hypothesis. The last module involves generation of a sentiment graph. It takes the classified reviews that are stored in the database as a input which are visually shown as the sentiment graph. The Figure 1 describes the functional architecture of the proposed system.
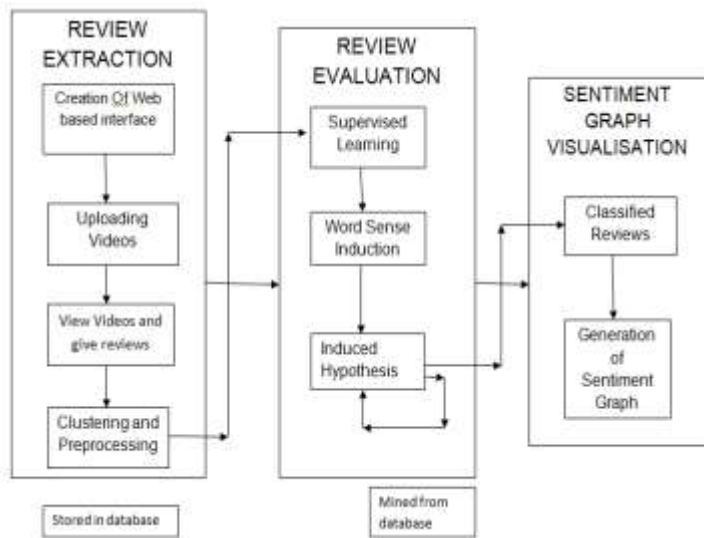
Fig.1 Proposed System

IV.IMPLEMENTATION

The implementation of the system involves the following modules:

A. Review Extraction

- Creation of Web based Interface.
- Extracting words from text using Word Clustering that influence the outcome of the result.
- A preprocessing of words are executed to avoid duplication in words.

Pseudocode: *Word Clustering*

```
Given: a dictionary D of size N
Create: table of clusters
i=0
while (D is not empty)
  G1=D[0]
  t=sizeof(D)
  for j from 1 to t
   if d1(G1,D[j])=1 then d1(G1,m1)=1<=i<=n D[j]->M1
   end for
remove G1 and all M1 members from D
(G1,M1)->C
i=i+1
end while
Return C
```

Pseudocode: *Preprocessing*

Sql query:

select tb1.a1,tb2.a2 from tb1,tb2
from tb2 inner join tb1 on tb1.id=tb2.id
where tb1.a1=""

B. Review Evaluation

- It involves refining text in form of relevant phrases, words, etc., which are contained in a set called hypothesis space.
- Induced Hypothesis is used for those words that are to be classified by the owner of the video.
- Supervised learning classifies the words into positive, negative and neutral which is implemented in coding part.
- Architecture for Support Vector Machine Learning is shown in Figure 2.
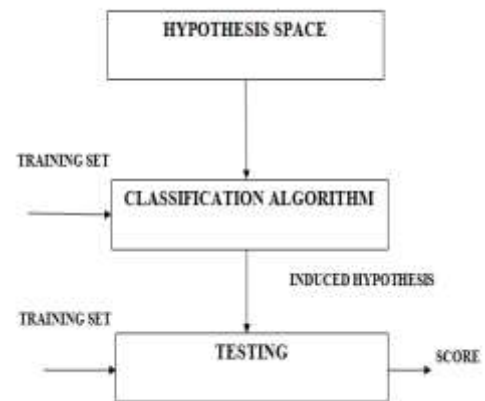

Fig.2 Support Vector Machine Learning

- A sample of reviews and their class to which it is classified into, is shown in the Table 1 for all types of sentiments.

Table1
Reviews and their class

Review Text	Class
Very Nice	Positive
Great!!!	Neutral
Very Bad	Negative
Average	Neutral
WOW!*	Neutral

C. Sentiment Graph Visualization

- It involves producing a sentiment graph for the classified reviews which are useful in a way of understanding the product.

Pseudocode:

```
video(){
 con.Open();
```

160

```
SqlCommand cmd = new SqlCommand("select
count(id) as vid,count(Status) as Status from reviews
where              vid='"              +
Convert.ToInt32(Session["id"].ToString()) + "' and
Status='Positive'", con);
 SqlDataAdapter dr = new SqlDataAdapter(cmd);
 DataTable dt = new DataTable();
 dr.Fill(dt);
 Chart.DataSource = dt;
 Chart.DataBind();
 con.Close();
```

V. RESULTS AND PERFORMANCE ANALYSIS

A. Classification of sentiment words

In the classification of the sentiment words, the reviews given by the user is classified as positive, negative and neutral by using Support Vector Machine Algorithm as shown in Figure 3.

Fig.3 Classified reviews

The process takes place by the reviews being parsed and traversed between database and the dictionary which will define the sentiment extraction patterns for the predicates mentioned.

The reviews given by the user may be sentence level, word or a document level and proper classification will be done and the score of the reviews are inferred with colors, where positive is represented by green, negative is represented by red and neutral is represented by yellow. The neutral words may have some hypothesis words present in it. These hypotheses are again classified using the Supervised Learning algorithm. This algorithm will classify the words recursively until proper score is obtained.

B. Sentiment Graph Visualisation for Sentiwords

A new approach called sentiment graph is introduced. This shows the classification of the reviews in the form of graph as shown in Figure 4
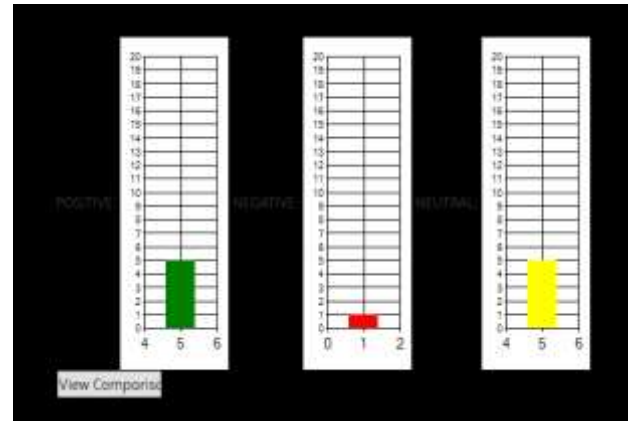

Fig.4 Sentiment Graph

The process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

Here the classification of reviews is represented using the bar graph. There exist a practical inefficiency wherein executive or users need to go through all the comments to find out the number of positive, negative, neutral comments before coming to a decision. The Sentiment graph reduces this difficulty.

The graph has X-axis and Y-axis, wherein the X-axis represents the sentiment field and Y-axis represents the number of reviews given by the user. This information will create a major revolution in marketing and advertising field.

C. Comparison of the Sentiment graph

The different videos uploaded by the admin are viewed by the users and then reviews given by them are compared for better understanding as shown in Figure 5
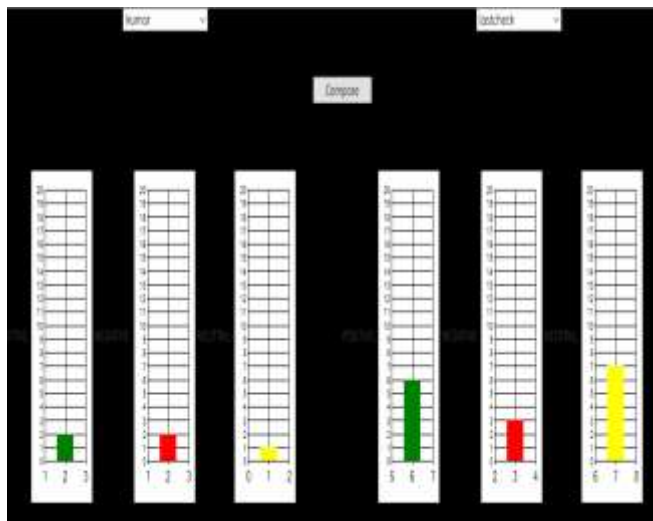
Fig.5 Comparison of Sentiment Graph

With the help of the comparison mentioned in the graph the release and the sales of the product can be determined. The accuracy of the classification is also improved in comparison. The clear picture of the sentiment classification is achieved here.

In this proposed system comparison between the two systems can be viewed by the user at the same time and proper perception and feedback for the videos uploaded by the admin is obtained. This the major result provided in this system which is not present in the existing system.

VI. CONCLUSION

Sentiment analysis is essential for anyone who is going to make a decision. Sentiment analysis is helpful in different field for calculating, identifying and expressing sentiment. It is helpful for everyone when they want to buy a product and they can decide which product is best. Sentiment analysis is very important for Enterprises and helps them to know what customers think about their products. Therefore companies can take decisions about their products based on customer's feedback Thus companies can modify their products features and introduce new products according to customer's opinion in a better and a faster way.

VII. FUTURE ENHANCEMENT

There is a lot of scope in analyzing the video and the images on the web. Nowadays with the advent of Facebook, Instagram and video vines, people are expressing their thoughts with pictures and videos along with text. Sentiment analysis will have to pace up with this change. Tools which are helping companies to change strategies based on Facebook

and Twitter will also have to accommodate the number of likes and re-tweets that the thought is generating on the social media. People follow and unfollow people and comments on social media but never comment. So there is a scope in analyzing these aspects of the web as well.

REFERENCES

[1]   Abbasi,A.H. Chen, and A. Salem, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," ACM Trans. Information Systems, vol. 26, no. 3, pp. 1-34, 2008.

[2]   Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Transactionson Knowledge and Data Engineering (TKDE), vol. 23, no. 3, pp.447-462, 2011.

[3]   Chenghua Lin, Yulan He, Richard Everson, Member, IEEE, and Stefan Ru¨ger, June 2012,"Weakly Supervised Joint Sentiment-Topic Detection from Text", IEEE transactions on knowledge and data engineering, VOL. 24, NO. 6

[4]   Dasgupta,S. and V. Ng, "Topic-Wise, Sentiment-Wise, or Otherwise? Identifying the Hidden Dimension for Unsupervised Text Classification," Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), pp. 580-589, 2009.

[5]   Liu,B. "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2012.

[6]   Pang,B. and LeeL,."Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-135,2008.

[7]   Rui Xia, FengXu, ChengqingZong, QianmuLi, Yong Qi, and Tao Li, August 2015," Dual Sentiment Analysis: Considering Two Sides of One Review", IEEE transactions on knowledge and data engineering, Vol.27, AnNo.8.