International Journal of Research in Advent Technology, Vol.2, No.2, February 2014

E-ISSN: 2321-9637

A survey on Sentiment Analysis of Texual Review

Hiteshkumar N Vegda[1], Prof.Tejal R Patel[2], Prof.Bhargesh B Patel[3]

Information Technology[1,2,3]

G.H.Patel College of Engineering and Technology,Vallabh Vidhyanagar, Gujarat, India[1,2,3]

hvegda@gmail.com[1], tejalrpatel@gcet.ac.in[2], bhargeshpatel@gcet.ac.in[3]

Abstract-Our day-to-day life has always been influenced by what people think. Ideas and opinions of others have always affected our own opinions. The explosion of Web has led to increased activity in Podcasting, Blogging, Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. As a result there has been an eruption of interest in people to mine these vast resources of data for opinions.\ Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. In this paper, we take a look at the various challenges and applications of Sentiment Analysis. We will look at Sources of data for sentiment analysis as well as tools available for sentiment analysis. We will also discuss various techniques to perform sentiment Analysis. And lastly the evaluation of sentiment classification.

Keywords- Machine learning, Opinion mining, Sentiment analysis (SA)

I. INTRODUCTION

1.1 *WHAT IS SENTIMENT ANALYSIS?*

Sentiment Analysis [1] is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall tonality of a document. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today. The Web is a huge repository of structured and unstructured data. The analysis of this data to extract latent public opinion and sentiment is a challenging task.

Liu et al. [1] defines a sentiment or opinion as a quintuple-

"<**oj**, **fjk**, **soijkl**, **hi**, **tl** >,

Where, **oj** is a target object,

fjk is a feature of the object oj,

soijkl is the sentiment value of the opinion of the opinion holder hi on feature fjk of object oj at time tl, soijkl is +ve,-ve, or neutral, or a more granular rating,

hi is an opinion holder,

tl is the time when the opinion is expressed.

The analysis of sentiments may be document based where the sentiment in the entire document is summarized as positive, negative or objective. It can be sentence based where individual sentences, bearing sentiments, in the text are classified. SA can be phrase based where the phrases in a sentence are classified according to polarity.

Sentiment Analysis identifies the phrases in a text that bears some sentiment. The author may speak about some objective facts or subjective opinions. It is necessary to distinguish between the two. SA finds the subject towards whom the sentiment is directed. A text may contain many entities but it is necessary to find the entity towards which the sentiment is directed. It identifies the polarity and degree of the sentiment. Sentiments are classified as objective (facts), positive (denotes a state of happiness, bliss or satisfaction on part of the writer) or negative (denotes a state of sorrow, dejection or disappointment on part of the writer). The sentiments can further be given a score based on their degree of positivity, negativity or objectivity.

1.2 *SENTIMENT ANALYSIS PROCESS*

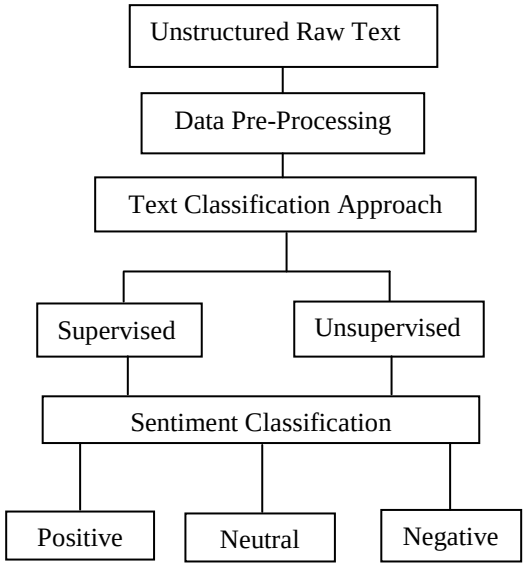

Figure1. Sentiment Analysis Process

The opinion mining process is explained in the figure1. The raw data is collected from various social media, Review Sites, Blogs, we can also write a crawler to extract the data from it. The Data set available online for research work.

After data collection we preprocess the data to have a structured set of reviews, Pre-processing the data is the process of cleaning and preparing the text for classification. The whole process involves several steps: online text cleaning,

white space removal, expanding abbreviation, stemming, stop words removal, negation handling and finally feature selection. so that we can apply classification techniques to classify the opinion either as positive, negative or neutral.

The supervised classification, Feature Extraction, It is the process where properties are extracted from the data, because the whole input data is too large to use in classification. The different feature extraction methods are N-gram model, Tf-idf (term frequency-inverse document frequency) Measure, Part-of-speech Tagger (POS). The Classification algorithm are nearest neighbour, naive Bayes, maximum entropy and Support vector machine (SVM) are applied. And in Unsupervised classification Point mutual information (PMI) is used.

II. APPLICATIONS OF SENTIMENT ANALYSIS

1. Applications to Review-related Websites Reviews and feedbacks [5] on almost everything, ranging from product reviews, to feedbacks on political issues are abundantly available over the Internet.Much like a search engine, a sentiment engine can be built to utilize this information.
2. Applications in Business Intelligence For many businesses [2], online opinion has turned into a kind of virtual currency that can make or break a product in the marketplace. This statement highlights the importance of sentiment analysis in businesses. The most obvious usage of Sentiment Analysis in business intelligence lies in understanding the user reviews to improve their products, and in turn, their reputation.
3. Opinion mining for Ad Placement In online systems that display ads as sidebars [2], it is helpful to detect WebPages that contain sensitive content inappropriate for ads placement; for more sophisticated systems, it could be useful to bring up product ads when relevant positive sentiments are detected, and perhaps more importantly, nix the ads when relevant negative statements are discovered.
4. Opinion mining for trend prediction Organization could perform trend prediction in sales using Opinion mining by tracking public viewpoints. In stock market we can analysis the sentiment related to detect whether the stock price will be higher or lower and help the investor to take decision related to buying or selling the stock [2].
5. Opinion mining for political domain Opinions matter a great deal in politics [2]. Sentiment analysis has specifically been proposed as a key enabling technology, allowing the automatic analysis of the opinions that people submit about pending policy or government-regulation, understanding what the voters is thinking, predicting the outcome of elections etc.
6. Summarization Opinion summarization [3] finds application when the number of online review of a product is large. This may make it hard for both the customer and the product manufactured. The consumer may not be able to read all the reviews and make an informed decision and the manufacturer may not be able to keep track of consumer opinion.
7. Voice of the Market (VOM) Voice of the Market is about determining what customers are feeling about products or services of competitors. Accurate and timely information from the Voice of the Market helps in gaining competitive advantage and new product development.
8. Products comparison It is a common practice for online merchants to ask their customers to review the products that they have purchased. With more and more people using the Web to express opinions, the number of reviews that a product receives grows rapidly. Most of the researches about these reviews were focused on automatically classifying the products into ‘‘recommended” or ‘‘not recommended” [11].

III. CHALLENGES FOR SENTIMENT ANALYSIS

Sentiment Analysis approaches aim to extract positive and negative sentiment bearing words from a text and classify the text as positive, negative or else objective if it cannot find any sentiment bearing words. In this respect, it can be thought of as a text categorization task. In text classification there are many classes corresponding to different topics whereas in Sentiment Analysis we have only 3 broad classes. Thus it seems Sentiment Analysis is easier than text classification which is not quite the case. The general challenges can be summarized as:

1. *Implicit Sentiment and Sarcasm*

A sentence may have an implicit sentiment even without the presence of any sentiment bearing words [3]. Consider the following examples.

“*How can anyone sit through this movie?”*

“*One should question the stability of mind of the writer who wrote this book.”*

Both the above sentences do not explicitly carry any negative sentiment bearing words although both are negative sentences. Thus identifying semantics is more important in Sentiment Analysis than syntax detection.

2. *Domain Dependency*

There are many words whose polarity changes from domain to domain [3].

Consider the following examples.

“*The story was unpredictable”*.

“*The steering of the car is unpredictable”*.

“*Go read the book”*.

In the first example, the sentiment conveyed is positive whereas the sentiment conveyed in the second is negative. The third example has a positive sentiment in the book domain but a negative sentiment in the movie domain (where the director is being asked to go and read the book).

3. *Thwarted Expectations*

Sometimes the author deliberately sets up context only to refute it at the end [3].

Consider the following example:
"This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up".
In spite of the presence of words that are positive in orientation the overall sentiment is negative because of the crucial last sentence, whereas in traditional text classification this would have been classified as positive as term frequency is more important there than term presence.

4. ***Pragmatics***

It is important to detect the pragmatics of user opinion which may change the sentiment thoroughly [2]. Consider the following examples:
"I just finished watching Barca DESTROY Ac Milan" "That final completely destroyed me".
Capitalization can be used with subtlety to denote sentiment. The first example denotes a positive sentiment whereas the second denotes a negative sentiment. There are many other ways of expressing pragmatism.

5. ***World Knowledge***

Often world knowledge needs to be incorporated in the system for detecting sentiments.
Consider the following examples:
"He is a Frankenstein".
"Just finished Doctor Zhivago for the first time and all I can say is Russia sucks".
The first sentence depicts a negative sentiment whereas the second one depicts a positive sentiment. But one has to know about Frankenstein and Doctor Zhivago to find out the sentiment.

6. ***Subjectivity Detection***

This is to differentiate between opinionated and non-opinionated text. This is used to enhance the performance of the system by including a subjectivity detection module to filter out objective facts. But this is often difficult to do [2].
Consider the following examples:
"I hate love stories".
"I do not like the movie "I hate stories"".
The first example presents an objective fact whereas the second example depicts the opinion about a particular movie.

7. ***Entity Identification***

A text or sentence may have multiple entities. It is extremely important to find out the entity towards which the opinion is directed.
Consider the following examples.
"Samsung is better than Nokia"
"Ram defeated Hari in football".
The examples are positive for *"Samsung"* and *"Ram"* respectively but negative for *"Nokia"* and *"Hari"*.

IV. SOURCES OF DATA FOR SENTIMENT ANALYSIS

A. *Review Sites*

A review site [3] is a website which allows users to post reviews which give a critical opinion about people, businesses, products, or services [3].

B. *Blogs*

The term web-log or blog [3], refers to a simple webpage consisting of brief paragraphs of opinion, information, personal diary entries, or links, called posts, arranged chronologically with the most recent first, in the style of an online journal.

C. *Micro-blogging*

Twitter is a popular micro blogging service where users create status messages called "tweets". These tweets sometimes express opinions about different topics. Twitter messages are also used as data source for classifying sentiment [3].

D. *Forums*

Forums [3] or message boards allow its members to hold conversations by posting on the site. Forums are generally dedicated to a topic and thus using forums as a database allows us to do sentiment analysis in a single domain.

E. *Social Networks*

Social networking [3] is online services or sites which try to emulate social relationships amongst people who know each other or share a common interest. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.

- Twitter- Twitter is an online social networking and micro blogging service that enables its users to send and read text-based posts of up to 140 characters, known as tweet.
- Facebook- Facebook is a social networking service and website launched in February 2004. The site allows users to create profiles for themselves, upload photographs and videos.

Social media is the new source of information on the Web. It connects the entire world and thus people can much more easily influence each other. The remarkable increase in the magnitude of information available calls for an automated approach to respond to shifts in sentiment and rising trends.

V. TOOLS AVAILABLES FOR SENTIMENT ANALYSIS

A variety of open-source text-analytics tools like natural language processing for information extraction and classification can be applied for sentiment analysis. The tools listed below can work on textual sources only [1].

LingPipe, OpenNLP, Stanford Parser and Part-of-Speech (POS) Tagger, NLTK, Opinion Finder, Tawlk/osae,

GATE, Textir, NLP Tool suite, Review Seer tool, Web Fountain, Red Opal, Opinion observer.

Along with these automated tools, there are various online tools like Twitrratr, Twendz,Social mention, and Sentimetrics are available to track the opinions in the web.

VI. TECHNIQUES FOR SENTIMENT ANALYSIS

All In order to give more imminent into the problem of opinion mining, in the following sections we discuss the domain overview and various types of opinion mining. The opinion mining is frequently associated with the topic information retrieval. The information retrieval algorithm works on factual data but the opinion mining works on subjective data. The task of opinion mining is to find the opinion of an object whether it is positive or negative and what features does it depict, and what features are appreciated, which are not etc. The notion of an opinion mining is given by Hu and Liu. They put most impact on their work and said that the basic components of an opinion are:

Opinion holder: it is the person that gives a specific opinion on an object.

Object: it is entity on which an opinion is expressed by user.

Opinion: it is a view, sentiment, or appraisal of an object done by user.

A. *Task of Opinion Mining at Document level*

Document level opinion mining [4] is about classifying the overall opinion presented by the authors in the entire document as positive, negative or neutral about a certain object. The assumption is taken at document level is that each document focus on single object and contains opinion from a single opinion holder. Apart from the document-level opinion mining, the next sub-section discusses the classification at the sentence level, which classify each sentence as a subjective or objective sentence and determine the positive or negative opinion.

B. *Task of opinion mining at Sentence level*

The sentence level opinion mining [4] is associated with two tasks. First one is to identify whether the given sentence is subjective (opinionated) or objective. The second one is to find opinion of an opinionated sentence as positive, negative or neutral. The assumption is taken at sentence level is that a sentence contain only one opinion for e.g., *"The picture quality of this camera is good."* However, it is not true in many cases like if we consider compound sentence for e.g., *"The picture quality of this camera is amazing and so is the battery life, but the viewfinder is too small for such a great camera"*, expresses both positive and negative opinions and we say it is a mixed opinion. For *"picture quality"* and *"battery life"*, the sentence is positive, but for *"viewfinder"*, it is negative. Like the document-level opinion mining, the sentence-level opinion mining does not consider about object features that have been commented in a sentence. For this the feature level opinion mining is discuss in the next sub-section.

C. *Task of Opinion mining at Feature level*

The task of opinion mining at feature level [4] is to extracting the features of the commented object and after that determine the opinion of the object i.e. positive or negative and then group the feature synonyms and produce the summary report. Liu used supervised pattern learning method to extract the object features for identification of opinion orientation. To identify the orientation of opinion he used lexicon based approach. This approach basically uses opinion words and phrase in a sentence to determine the opinion. The working of lexicon based approach is described in following steps.

- Identification of opinion words
- Role of Negation words
- But-clauses

In contrast , Hu and Liu do customer review analysis through opinion mining based on feature frequency, in which the most frequent features is accepted by processing many reviews that are taken during summary generation. In opposite to Hu and Liu, Popescu and Etzioni, improved the frequency based approach by introducing the part-of relationship and remove the frequent occurring of noun phrases that may not be features.

D. *Opinion Mining in Compound sentence*

In this sub-section the following methodology we use to determine the opinion in compound sentence of a movie review domain [4]:

- *Sentence classification*

In the sentence classification we go to individual compound sentences to determine whether a sentence is subjective or express an opinion and if so, whether the opinion is positive or negative (called sentence-level sentiment classification).

For example, *'Desi Boyz'* - highly entertaining comedy gives the positive opinion and *'Damadamm'* clearly has no Dum gives the negative opinion. The following activities are done within sentence classification:

- *Splitting of the document into sentences*

Given a document about the movie reviews, the document is segmented into individual sentences by the help of sentence delimiter. Here problem is that most of the reviews are found on movie forums or blog sites where normal users post their opinions in their informal language which do not follow strict grammatical rules and punctuations. The identification of full stop in the sentence does not mark the end of sentence sometimes. Such as date 12.1.2012, movie short forms T.M.K expressed in reviews, hence we have to use rule based pattern matching to identify sentence boundary. Second problem is that people generally use internet slang words like *OMG, cuteeeee* etc. e.g. actress is cuteeee. Here there is not such word in dictionary like *cuteeee* but it refers to cute. We will do N gram matching of such words with pre compiled dictionary of movie related words.

- *Determining whether the sentence is opinionated*

We will use boot strap approach proposed by Riloff and Wiebe for the task of subjective sentences identification. It will use a high precision (and low recall) classifiers to extract a number of subjective sentences collected from various movie review sites. From this subjective sentence a set of patterns will be learned. The learned patterns will be used to extract more subjective and objective sentence. The subjective classifier will look for the presence of words from the pre compiled list, while the objective classifier tries to locate sentences without those words. In the example 1 all sentences except "*M Gud reviews about film released on 12.1.2012"* are opinionated.

- *Determining whether the opinionated sentence is compound sentence*

A compound sentence is a sentence that contains two or more complete ideas (called clauses) that are related. These two or more clauses are usually connected in a compound sentence by a conjunction. The coordinating conjunctions are *"and", "but", "for", "or", "nor", "yet", or "so".* We will use plain pattern matching to find out the presence of coordinating conjunctions. If they are present in the given sentence then it will be identified as compound sentence. Output of this will be set of compound sentences

- *Opinion Mining in Compound sentence*

In this section we focus on opinion expressions in the compound sentence of a movie review that gives the opinion on the individual feature of the movie and the opinion of the movie as whole i.e. positive or negative sentiments. Apart from this we also determine the sentiment score towards various features of a movie, such as cast, director, story and music. Sentiment scores are used to classify the sentiment polarity (i.e. Positive, negative or neutral) of clauses or sentences. The linguist approach makes use of both a domain-specific lexicon (specify the noun related terms like actor, director etc.) and a generic opinion lexicon(specify the property of movie related terms),derived from SentiWordNet to assign a prior sentiment score to each word in a sentence. For example,

He says, "The film T.M.K's story is filled with a great plot, the actors are first grade, and actress is cuteeee. The supporting cast is good as well, but, movie can't hold up".

The compound sentences (1 and 2) are divided into *"The film T.M.K's story is filled with a great plot" "the actors are first grade" "actress is cute"; "The supporting cast is good as well" and "movie can't hold up"*. After dividing the compound sentences into separate sentences or clauses, a sentiment score toward each movie features (e.g. Story, actors, actress, supporting cast or the film features) is calculated. (For example like the sentiment score is from 0 to 5 here 0 indicates the most negative opinion, 5 indicate the most positive opinion and between 0 to 5 we have to make the rules for positive opinion and negative opinion.) After calculating the sentiment score for each clause or sentence, the sentiment score for each review features and the overall sentiment score for the whole sentence is calculated.

VII. TECHNIQUES FOR SENTIMENT CLASSIFICATION

The literature survey done indicates two types of techniques include machine learning and semantic orientation.

A. *Machine Learning*

Several Machine Learning methods have been studied [12]. Prominent methods are: Naive Bayes Classification, Maximum Entropy Classification, and Support Vector Machines. In his work, Pang Lee et al., compared the performance of Naïve Bayes, Maximum Entropy and Support Vector Machines in SA on different features like considering only unigrams, bigrams, combination of both, incorporating parts of speech and position information, taking only adjectives etc. The result has been summarized in the Table 1. It is observed from the results that:

- Feature presence is more important than feature frequency.
- Using Bigrams the accuracy actually falls.
- Accuracy improves if all the frequently occurring words from all parts of speech are taken, not only Adjectives.
- Incorporating position information increases accuracy.
- When the feature space is small, Naïve Bayes performs better than SVM.

But SVM's perform better when feature space is increased.

When feature space is increased, Maximum Entropy may perform better than Naïve Bayes but it may also suffer from overfitting. Despite its simplicity and the fact that its conditional independence assumption clearly does not hold in real-world situations, Naïve Bayes-based text categorization still tends to perform surprisingly well; indeed, Domingos and Pazzani show that Naive Bayes is optimal for certain problem classes with highly dependent features. Maximum Entropy Classification (Max Ent, or ME, for short) is an alternative technique which has proven effective in a number of natural language processing applications. Nigam et al. show that it sometimes, but not always, outperforms Naive Bayes at standard text classification.

Table 1 - Comparison of Accuracies using Naive Bayes, Maximum Entropy and Support Vector Machines [12].

Features	**No. of features**	**Frequency or Presence**	**NB**	**ME**	**SVM**
Unigrams	16165	Freq.	78.7	N/A	72.8
Unigrams	16165	Pres.	81.0	80.4	82.9
Unigrams	32330	Pres.	80.6	80.8	82.7

+Bigrams					
Bigrams	16165	Pres.	77.3	77.4	77.1
Unigrams + POS	16695	Pres.	81.5	80.4	81.9
Adjectives	2633	Pres.	77.0	77.7	75.1
Total 2633 unigrams	2633	Pres.	80.3	81.0	81.4
Unigrams + position	22430	Pres.	81.0	80.1	81.6

B. *Sementic Orientation*

Problem of Opinion mining can be categorized as sentiment classification and feature based opinion mining. Problem of Opinion mining can be categorized as sentiment classification and feature based opinion mining [12].

Classification of Approaches of Sentiment Orientation

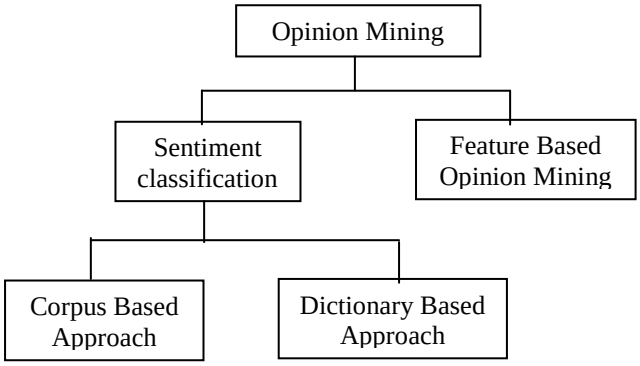

Figure2. Granularity Levels of Sentiment analysis [12]

- *Corpus Based Approach*

Popular corpus-driven method is to determine the emotional affinity of words which is to learn their probabilistic affective scores from large corpora. Mihalcea and Liu have used this method to assign a happiness factor to words depending on the frequency of their occurrences in happy-labeled blog posts compared to their total frequency in a corpus containing blog posts labeled with “happy” and “sad” mood annotations. They also compare the happiness factor scores of words with the scores in the ANEW list.

- *Dictionary Based Approach*

These approaches have used lexical resources such as WordNet to automatically acquire emotion-related words for emotion classification experiments. Starting from a set of primary emotion adjectives, Alm et al. retrieve similar words from WordNet utilizing all senses of all words in the synsets that contain the emotion adjectives. Researchers also exploit the synonym and hyponym relations in WordNet to manually find words similar to nominal emotion words.

C. *Features for Opinion Mining*

Feature engineering is an extremely basic and essential task for Opinion Mining. Converting a piece of text to a feature vector is the basic step in any data driven approach to Opinion. Some commonly used features in Opinion Mining and their critiques have been discussed in the following sections.

- *Term Presence vs. Term Frequency*

Term frequency [12] has always been considered essential in traditional Information Retrieval and Text Classification tasks. But it is found that term presence is more important to Sentiment analysis than term frequency. That is, binary-valued feature vectors in which the entries merely indicate whether a term occurs (value 1) or not (value 0).

- *Term Position*

Words appearing in certain positions in the text carry more sentiment or weightage than words appearing elsewhere. This is similar to IR where words appearing in topic Titles, Subtitles or Abstracts etc are given more weightage than those appearing in the body.

- *N-gram Features*

The n-gram [2] is a contiguous sequence of n items from a textual or spoken source In case of unigram (n=1), each text is a document and is spilt up into words. The term frequency (tf(w,d)) is the number of times that a word w occurs in a document d. The term presence tp(w,d) only checks if a word w is present within a document d which result in binary value. In case of n=2(big rams) the items consist of two consecutive words i.e. the set contains all combination of two words that are consecutive in the original text.

- *Tf-idf Measure*

The tf-idf (term frequency-inverse document frequency)[2] measure is a statistic that reflects the importance of a word across a set of document. The inverse document frequency is used to measure the rareness of a word across the entire document. Higher the value of inverse document frequency rare the word across the set of document.

- *Parts of Speech*

Parts of Speech information is most commonly exploited in all NLP tasks. One of the most important reasons is that they provide a crude form of word sense disambiguation. In English language POS [2] examples are noun, verb, adverb and adjective.

- *Adjectives only*

Adjectives [12] have been used most frequently as features amongst all parts of speech. A strong correlation between adjectives and subjectivity has been found. Although all the parts of speech are important people most commonly used adjectives to depict most of the sentiments and a high accuracy have been reported by all the works concentrating on only adjectives for feature generation.

VIII. EVALUATION OF SENTIMENT CLASSIFICATION

In general, the performance of sentiment classification is evaluated by using four indexes. They are Accuracy, Precision, Recall and F1-score. The common way for computing these indexes is based on the confusion matrix as shown below:

Table 2 - Confusion Matrix [1]

	Predicted positives	Predicted negatives
Actual positive Instances	Number of True Positive instances (TP)	Number of False Negative instances (FN)
Actual negative Instances	Number of False Positive instances (FP)	Number of True Negative instances (TN)

These indexes can be defined by the following equations [1]:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1} = \frac{2 \times \text{Pr}ecision \times \text{Re}call}{\text{Pr}ecision + \text{Re}call}$$

Accuracy is the portion of all true predicted instances against all predicted instances. An accuracy of 100% means that the predicted instances are exactly the same as the actual instances. Precision is the portion of true positive predicted instances against all positive predicted instances. Recall is the portion of true positive predicted instances against all actual positive instances. F1 is a harmonic average of precision and recall.

IX. CONCLUSION AND FUTURE WORK

This paper introduced and surveyed the field of sentiment analysis and opinion mining. It tried to showcase from basics, different techniques, evaluation method, and wide range of applications that are commonly used for Sentiment Analysis. It has been a very active research area in recent years. In fact, it has spread from computer science to management science. Sentiment analysis can be applied to a wide domain to classifying and summarizing review and prediction. Finally, this paper concludes saying that all the sentiment analysis tasks are very challenging.

In future, more work is needed on further improving the performance measures. Sentiment Analysis can be applied for many new applications. The techniques and algorithms used for Sentiment Analysis are very fast, and many of the study remain unsolved. The main challenging aspects exist in the use of Sentiment Analysis classifications, dealing with negation expressions; produce a summary of opinions based on product features, complexity of sentence/document, handling of hidden product features, etc. Many future research works could be committed to these challenges.

References

[1] S Padmaja and Prof. S Sameen Fatima, "Opinion mining and sentiment analysis – an assessment of peoples' belief: a survey", International journal of ad hoc, sensor & ubiquitous computing (IJASUC) vol.4, no.1, Feb 2013.

[2] Mrs. Vijyalaxmi m, Mrs. Shalu Chopra, Mrs. Sangeeta Oswal, Mrs. Deepshikha chaturvedi, "The How, When and Why of Sentiment Analysis", International journal of computer technology& application, vol 4 (4), 660-665, 2013.

[3] Akshi kumar, Aeeja mary sebastian, "Sentiment Analysis: a perspective on its past present and future", I.J Intelligent systems and applications, ecs,page 1-14,2012.

[4] Nidhi Nishra, C.K.Jha, "Classification of Opinion Mining Techniques", International Journal of Computer Applications, Volume 56– no.13, Oct 2012.

[5] V. s. Jagtap, Karishma Pawar, "Analysis of Different Approaches to Sentence-level Sentiment Classification", International Journal of Scientific Engineering and Technology, Volume 2, Issue 3, April 2012.

[6] Pravesh kumar singh, Mohd Shahid husain, "Analytical study of Feature Extraction Techniques in Opinion Mining", International Conference on Advance in Computing and Information Technology, page 85-94, 2013.

[7] S.l. Ting, W.h. Ip, Albert H.c. Tsang, "Is Naïve Bayes a Good Classifier for Document Classification?", International Journal of Software Engineering and Its Applications vol. 5, no. 3, July, 2011.

[8] Emma Haddi, Xiaohui Liu, Yong Shi, "The Role of Text Pre-Processing Sentiment Analysis", Information Technology and Quantitative Management, Procedia Computer Science 26 -32, Elsevier 2013.

[9] P. Waila, Marisha, V.k. Singh, M.k. Singh, "Evaluating Machine Learning and Unsupervised Semantic Orientation Approaches for Sentiment Analysis of Textual Reviews", IEEE International Conference on Computational Intelligence and Computing Research, 2012.

[10] Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah Khan, "A Review of Machine Learning Algorithms for Text-Documents Classification", Journal of Advances In Information Technology, VOL. 1, February 2010.

[11] Huifeng Tang, Songbo Tan , Xueqi Cheng, "A Survey on Sentiment Detection of Reviews ", Expert Systems with Applications, page 10760 - 10773, Elsevier 2009.

[12] Nilesh m. Shelke, shriniwas Deshpande, Vilas Thakre, "Survey of Techniques for Opinion Mining" , international Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.

[13] Mr. Saifee Vohra, Prof. Jay Teraiya, "Applications and Challenges for Sentiment Analysis: A Survey", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2, February- 2013.