A Survey on Sentiment Analysis and Opinion Mining: A need for an Organization and Requirement of a customer

Ravendra Ratan Singh Jandail#1, Pradeep Sharma#2, Chetan Agrawal#3
Department of Computing Science & Engineering, Galgotias University Greater Noida, India
ravendraratansingh@yahoo.co.in, pradeep.sharma@galgotiasuniversity.edu.in

Abstract - Sentiment analysis and Opinion mining is the computational study of User opinion to analyze the social, psychological, philosophical, behavior and perception of an individual person or a group of people about a product, policy, services and specific situations using Machine learning technique. Machine learning for text analysis technically has always been very challenging as its main goal is to make computers able to learn and automatically generate emotions like a human as it is practically very useful in real life scenarios. After a boom in web 2.0 technology this field became the most interesting area for researchers because the social media has grown as the fastest medium for availability of opinions. There are many commercial tools available in the market and many researchers have proposed their solutions for opinion extraction, but still there are some problems of text classification and sentiment extraction in opinion mining. These problems arise due to different behaviors, manners and textual habits of users. A sentence can be positive for one, but it may have a negative impact on other so it will be a problem for a machine to generate its emotion. A negative sentence can be written in a positive manner like “What a great camera! It consumes more battery power, this sentence has a negative opinion about a camera, but it consists only positive keywords. There are mainly four predominating problems viz. subjectivity classification, word sentiment classification, document sentiment classification and opinion extraction. Data mining algorithms are easy to implement, but concludes to poor accuracy meanwhile the machine learning technique provides better accuracy, but requires a lot of training time, so there should be a hybrid technique which has the advantages of both the techniques. This survey focuses on various Data Mining algorithms, Machine learning techniques and a brief review stating the comparative analysis of these algorithms. We have followed a systematic literature review process to conduct this survey and also mentioned the future aspects of sentiment analysis and opinion mining.

Keyword – Sentiment Analysis, Opinion Analysis, Market Research, machine Learning

I. INTRODUCTION

Sentiment analysis and opinion mining is the process of analyzing the market research on user opinion, which is need of an organization and requirement of a customer. Market research is a continuous process for gathering data on product characteristics, supplier’s capabilities and the business practices that surround them, plus the analysis of that data to make acquisition Decisions[1]. Opinion Mining for Market research require a large amount of useful data which consist of user reviews about the characteristics and behavior of a particular product for decision making. Due to the advancement in technology, today’s web technology and social media are the best way to collect user’s opinion for any particular product perception for Sentiment analysis. Millions of reviews of various products are posted daily on twitter, Facebook, MySpace, YouTube, Flickr, IM eem, Last FM, Digg, Bebo, Click, hi5, LinkedIn, Live Journal, tumbler and other social networking websites. There are many websites which are collecting reviews and provide information to the organization. The information provided by some product specific websites (like www.gsmarena.com, www.amazon.com) is a rich and useful source for marketing researchers, social psychologists, organization and others interested in extracting and mining opinions, views, moods, and attitudes [2]. Sentiment analysis uses this information to extract knowledge for decision making about the market and a particular product that, what is the actual requirement of a customer? Not, what the company or organization is providing. In decision making for a specific Product perception the sentiment analysis provides 1). Market analysis, whether the market is positive or negative for specific product 2). What is the actual requirement of a customer? 3). Who is the user? (Information about the user, for example – what will be the most affordable prices in Bhutan for Apple iPhone). 4). How many types of users are present in a particular country or Region or Market? (Identifying the need of product example- this is the time of multimedia mobile phone, but still few people are demanding cheap and simple mobile phone but only Nokia mobile phone). 5). Product reliability 6). Customer faith and manufacturer responsibility 7). Cause of business success and failure

Tamizhan College of Engineering and Technology (ISO 9001:2008 Certified Institution), Tamilnadu, INDIA. Page 17
II. LITERATURE REVIEW

Sentiment analysis, a sub-discipline within data mining and computational linguistics, refers to the computational methodology for mining, understanding and assessing the opinions expressed in many opinion-rich resources like blogs, discussion forums [3]. In industries it is known as Sentiment analysis while in academics it is preferred as sentiment analysis and opinion mining both. Sentiment analysis is a part of text mining. The term sentiment analysis was first used by Jeonghee Yi and Na constructing in 2003 in their paper Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques [4]. Human opinion are always be considered for decision making. Before technological development human opinion were collected by paper survey and word of mouth but now days large amount of human perception are available in the textual form on the web sites. The first textual data processing system “A business intelligence system” was proposed by H.P. Luhn in 1958 [5]. Text mining is also used in industries as information retrieval. Information retrieval in the broader sense deals with the entire range of information processing, from data retrieval to knowledge retrieval, although, information retrieval is a relatively old research area where first attempts for automatic indexing where made in 1975 [SWY75], it gained increased attention with the rise of the World Wide Web and the need for sophisticated search engines [7]. In 1969 George Gerbner, professor of Mass communication established violence Profile for different TV networks see how various groups portrayed in his article “LIVING WITH TELEVISION: THE DYNAMICS OF THE CULTIVATION PROCESS” [8]. Analysis of kinship terminology (Goodenough, 1972) In literary and rhetorical analysis, thematic analysis (Smith 1972; Ile 1982, 1989), subjectivity analysis (Uspensky 1973, Banfield 1982). Determine author is positive, neutral or negative toward some part of a document (Hearst, 1992) [9]. The importance of word-of-mouth has been widely documented in the existing literature (Anderson, 1998; Goldenberg, Libai, & Muller, 2001; Zhu & Zhang, 2006; Stokes & Lomax, 2002) [10]. A classic example of subjectivity analysis is the detection of opinionated text in order to distinguish it from an objective (Finn, 2003; Ng et al., 2006; Ni et al., 2007). Pang et al. (2002) employ three machine learning methods to classify movie reviews into two Classes: positive and negative, neutral review were also considered in this study which is useful in result analysis. Pang and lee (2005) applied M et al algorithm to solve the problem which alters the multi-class problem. Kim and hovy (2005) also work on classifying election poll. As the use of internet is increasing User are posting their opinion on the micro blogs on the social media so the twitter data was also analyzed by Leskovec et al. (2009) Go et al. (2009), (Bermingham and Smeaton, 2010). Barbosa and Feng (2010) also developed a model to analyze the thousands of tweets to give a prediction about next thousand tweets. M icro Blogs are a broadcast medium in the form of blogging. Blogs Consist of personal opinion made by the author or blog owner and displayed in chronological order. M icro blogs differ from traditional blogs through their smaller size of published elements. This Medium allows people to exchange small elements of content such as short sentences, or links to Videos, images and other media (Kaplan and Haenlein, 2011). Since 2010 the twitter is most wildly using for sentiment Analysis because a tweet consists only 140 letters. To information retrieval from twitter. Abel, Celik and Siehndel (2011) proposed an algorithm, it also retrieves information from linked websites giving by a user in his tweets. One application of this concept in information retrieval is the well-known PageRank algorithm used by Google, which is based on the idea that the relevance of a website depends on the importance of the relevant websites linked to it [12]. Cotelo, Cruz and Troyano (2012) proposed an algorithm to dynamically update the search keywords in twitter. Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N.(2013) designed an ontology based expert system for twitter sentiment analysis. Ge M and Jiebo Luo developed an algorithm for automatically analyzing the sentiment of facial expressions in social Photos, especially those of politicians in the context of elections in his article “ IS A PICTURE WORTH 1000 VOTES? ANALYZING THE SENTIMENT OF ELECTION RELATED SOCIAL PHOTOS” An intelligent system for sentiment extraction from natural audio stream was also developed using maximum entropy models and automatic speech recognition system [13]. So on the way of development from 1960 to 2013 various researches were implemented for sentiment analysis using M achine intelligence. Since early 2000, sentiment analysis has grown to be one of the most active research areas in natural language processing [6].
Management. The true usability of a sentiment system is revealed when you are able to drill into each “piece of the sentiment pie” or a point on the sentiment trend graph. Prioritize results by relative importance, reach out when necessary, and have each interaction become part of that conversation record. Its goal is to enrich each conversation and deepen engagement with the same person across multiple channels. The Holy Grail of Social CRM is the ability to close the analytical loop between publicly expressed sentiment, engagement action, subsequent purchase intent, and ultimately, product purchase.

IV. SENTIMENT ANALYSIS REQUIREMENT OF A CUSTOMER

“Water water everywhere and not a drop to drink” [41], this sentence is summarizing everything that what exactly the customer wants. If I elaborate this sentence suppose a person is patient of malaria fever, but the medical store have every kind of medicine except Malarial then what is the use of other such prescription? Which can’t cure the person? Sentiment Analysis and opinion mining will be helpful to resolve such kind of issue and to understand the actual customer requirement? Opinion aggregation is necessary requirement at the end user’s perspective. The number of online reviews that a product receives grows rapidly, especially for some popular products. Furthermore, many reviews are long and have only a few sentences containing opinions on the product. This makes it hard for a potential customer to read them to make an informed decision on whether to purchase the product[2] 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 people like it or what they complain about. So opinion mining is helpful to customer in decision making weather a customer should go for this product, policy and services.

V. SENTIMENT ANALYSIS IN MARKET RESEARCH

Market research is a continuous process for gathering data on product characteristics, supplier’s capabilities and the business practices that surround them, plus the analysis of that data to make acquisition decisions [14] for organization and customer. The market research paradigm includes the data collection from the market and analyze the collected data for a particular product. The sentiment analysis uses this information to produce results about customer satisfaction and product reliability, product quality and other market demanding feature in the current commercial scenarios. Sentiment analysis is a Commercial practice regarding customizing, modifying, up dating of product or tailoring services, buyer interest, previous buyer feedback, buyers financial Condition, effectiveness of advertising and discounts under which commercial sales are made. Sentiment analysis provides a faster, simpler and less expensive alternative to traditional qualitative market research techniques like observations, interviews and even ethnography as well as provided information in real time [16][23]. The demand of market research is increases with the increasing the literacy rate of user and competition in the market with the other companies. Every organization has its research and development center where the researcher is working with experienced marketing strategist on the other hand some organization outsource this work to other vendors. The following figure shows the relationship that how the sentiment Analysis is giving its importance in the Market and

A. Principal of Market Research

The following six principal were proposed in Market Research A. Start early B. Involve users, C. Communication between consumer and supplier D. Think of market research as an iterative process E. Tailor the investigation F. Refine as you proceed [15].

B. Market Research Approaches

Research is carried out in two basic ways Qualitative Research Approach - Qualitative research involves finding people’s opinion, how much they are satisfied and, Customer faith, what they say, they think and how they say they feel. This approach provides this subjectivity information about the product. Quantitative research focuses on measuring an objective fact. Key to conducting quantitative research is definition of variables of interest and to a large extent a sense of detachment in the data collected by the researcher. Quantitative research analyzes data using statistics and relies on large samples to make generalized statements [24][16]. Various research methodologies have been applied using data mining, machine learning technique and natural language processing technique to analyze the market research. Various algorithms were developed for specific product review analysis (wu, wang and yi 2013), movie review analysis, Cell phone review analysis, Stock market price analysis(Aparna Anant Bhat, Sowmya Kamath S. 2013), urgent decision making(Qiansheng Zhang, Fuchun Liu, Yirong Huang, October 2013)[17]. Martínez, García & Sánchez develop a financial ontology algorithm for semantic based sentiment analysis in financial news [18]. For agricultural importance in Japan Okada, Hirokawa and Hashimoto did a survey of farmer opinion and analyze to increase the production of farm products (2013) [19][23].
VI. LEVEL OF SENTIMENT ANALYSIS

A. Phrase level sentiment analysis

The phrases that contain opinion words are found out and a phrase level classification is done. In some cases, the exact opinion about an entity can be correctly extracted but in some other cases, where contextual polarity also matters, the result may not be fully accurate [20].

B. Sentence level Sentiment Analysis

In the sentence level sentiment analysis, the polarity of each sentence is calculated. The subjective sentences contain opinion words which help in determining the sentiment about the entity. After which the polarity classification is done into positive and negative classes [20].

C. Document Level Sentiment Analysis

In this document level classification, a single review about a single topic is considered. But in the case of forums or blogs, comparative sentences may appear. Both supervised and unsupervised learning methods can be used for the document level classification [20].

VII. SENTIMENT DETECTION

Sentence Analysis Consist Five continuous, systematic steps from Searching user opinion to reporting results. a). Searching review from the various search engine b). Collecting Important review related to product policy and services c). Information Extraction d). Result Analysis e). Reporting Result

A. Subjectivity Classification

Subjectivity in natural language refers to aspects of language used to express opinions and evaluations (Wiebe, 1994). The Problem of Subjectivity Classification is to distinguish the relevant text from the topic of the irrelevant text. For example if a person did five tweets about Barack Obama and Man Mohan Singh then Subjectivity Classification Distinguish the Barack Obama tweets from the Man Mohan Singh Tweets. It can be stated as

T (total tweets) = {B (Barack Obama tweets) M (Man Mohan Singh Tweets)}.

B.Sentiment classification

Sentiment classification includes two kinds of classification forms, i.e., binary sentiment classification and multi-class sentiment classification. Given a document set D = {d1, . . . , dn}, and a predefined categories set C = {positive, negative}, binary sentiment classification is to classify each di in D, with a label expressed in C. If we set C* = {strong positive, positive, neutral, negative, strong negative} and classify each di in D with a label in C, the problem changes to multi-class sentiment classification [2].

VIII. DATA MINING AND MACHINE LEARNING TECHNIQUES

A. Data Mining Technique-

Sentiment analysis is a part of Data Mining. Data mining is a process of discovering an interesting pattern and knowledge from large amounts of unstructured data [21]. The main motive to apply the data mining method on text document is to structure the data, which can be analyzed easily and helpful in extracting the required knowledge within the accuracy as much as possible. In Sentiment analysis, classification, clustering, information extraction and explorative text mining: visualization methods widely used [22].

1. Classification

Classification is supervised learning technique and a common fundamental technology in many current opinion-mining and Sentiment-analysis applications is classification. Classification is the process of finding a process (or function) that describe and distinguish data class or concept. This model is designed by analyzing set of training data. In the case of classification, we are given a database of objects that are labeled with predefined categories or classes. We are required to learn from these objects in a model that separates them into the predefined categories or classes. Then, given a new object, we apply the learned model to assign this new object to one of the classes. In the more general situation of regression, instead of predicting classes, we have to predict real-valued fields [22][25][26].

2. Clustering

A clustering method was proposed to map implicit aspect expressions, which were assumed to be sentiment words, to their corresponding explicit aspects this is also called unsupervised learning. Here, we are given a database of objects that are usually without any predefined categories or classes. We are required to partition the objects into subsets or groups such that elements of a group share a common set of properties. Moreover the partition should be such that the similarity between members of the same group is high and the similarity between members of different groups is low [6][22][26].

3. Association

Class association rules are a special type of association rules with a fixed class attribute. Association analysis is useful for discovering interesting relationship hidden in large data sets and uncovered relationship can be represented in the form of association rule or set of frequent items [27]. An association rule is a pair that we write as X \rightarrow Y, where X and Y are two item sets and X \cap Y = NULL the item set X is called the antecedent of the rule. The item set Y is called the consequent of the rule [6][25].
Machine Learning Techniques

1. Sentiment Classification Using Supervised Machine Learning

Sentiment classification is usually formulated as a two-class classification problem, positive and negative. Training and testing data sets are usually product reviews. Sentiment classification is essentially a text classification problem. Traditional text classification mainly classifies documents of different topics. In text classification problem, any existing supervised learning method can be applied, e.g., naïve Bayes classification, and support vector machine (SVM). max entropy, Decision tree, Neural Network etc. Pang, Lee and Vaithyanathan (2002) was the first paper to take this approach to classify movie reviews into two classes, positive and negative. Like other supervised machine learning applications, the key to sentiment classification is the engineering of a set of effective features. Some of the example features are-
1) Terms and their frequency 2) Part of speech 3) Sentiment words and phrases 4) Rules of opinions 5) Sentiment shifters 6) Syntactic dependency[6]

a. Naïve Bayes model for sentiment classification

Naïve Bayes classifier is a commonly used supervised machine learning algorithm. This approach presupposes all sentences in opinion or factual articles as opinion or fact sentences. Recent years have seen a rapid growth in non-topical text analysis, in which characterizations are sought of the opinions, feelings, and attitudes expressed in a text, rather than just the subjects[2][39]. Naïve Bayesian classifier based on Bays theorem. Naïve Bayesian Classification assumes, that the effect of an attribute value of a given class is independent of the value of the other attributes. This assumption is called class conditional independence. It can include the counts of positive and Negative words in the sentence, as well as counts of the polarities of sequences of semantically oriented words.it also includes parts of speech tagging with their negative and positive polarity.

Naïve Bayes assigns a document dj represented by a vector dj* to the class Ci that maximize P(Ci/dj*) by applying Bayes rules as follows P(Ci/dj*) = P(dj*/Ci)P(Ci)/P(dj*)

Where P(dj*) is the probability that a randomly picked document d has vector dj* as its representative, and P(c) is the probability that a randomly picked document belongs to class c. To estimate the term P(dj*/c) Naïve Bayes decomposes it by assuming all the features in dj* are conditionally independent i.e.[2][10].

P(Ci/dj*) = \frac{P(Ci)P(dj*/Ci)}{P(dj*)}

b. Support Vector Machine for sentiment classification

Support Vector Machines (SVM) are primarily Classifiers that can classify by constructing hyper planes that separate cases that belong to different categories. In our application, the SVM classifier should first be trained on a set of user reviews so that the machine gains knowledge for effective categorization. A Support Vector Machines (SVM) is a supervised classification algorithm that recently has been applied successfully to text classification tasks. As usual a document d is represented by a possibly weighted vector (td1: : : ; tdN) of the counts of its words. A single SVM can only separate two classes a positive class L1 (indicated by y = +1) and a negative class L2 (indicated by y = -1). In the space of input vectors a hyperplane may be defined by setting y = 0 in the following linear

\[ y = w^T x + b = 0 \]

Fig-2 Hyperplane with maximal distance (margin) to examples of positive and negative classes constructed by the support vector machine. The SVM algorithm determines a hyperplane which is located between the positive and negative examples of the training set. The parameters of are adapted in such a way that the distance is called margin between the hyperplane and the closest positive and negative example documents is maximized [35][2]. SVMs can be used with non-linear predictors by transforming the usual input features in a non-linear way, e.g. by defining a feature map

\[ \phi(t_1, \ldots , t_N) = [t_1, \ldots , t_N, t_1^2, t_1t_2, \ldots , t_{N-1}^2, t_N^2] \]

[35][2].

c. Nearest Neighbor Classifier

The method K-nearest neighbor classification was described by early 1950. Instead of building explicit models for the different classes we may select documents from the training set which are “similar” to the target document. The class of the target document subsequently may be inferred from the class labels of these similar documents. If k similar documents are considered, the approach is also known as k-nearest neighbor classification. The Nearest Neighbor classifier is based on learning by analogy, that is by comparing a given test tuple that is similar to it. The Training tuples are described by an attribute. Each tuple represents a point in a n-dimensional space.
When a given unknown tuple a k nearest neighbor classifier search, the pattern space for the k training tuple that are closed to the unknown tuple. These k training tuples are the k “nearest neighbor” of the unknown tuple. [35][40]

4. Regression-Based Classifiers

Regression modeling is a method which is commonly used in order to learn the relationships between real-valued attributes. Typically, these methods are designed for real valued attributes, as opposed to binary attributes. This is however not an impediment to its use in classification, because the binary value of a class may be treated as a rudimentary special case of a real value, and some regression methods such as logistic regression can also naturally model discrete response variables[30].

5. Decision Trees

During the late 1970 and early 1980s, J. ross Quinlan a researcher in machine learning developed a decision tree algorithm known as ID3 (iterative Dichotomiser). This work expanded on earlier work on concept learning system described by E.B. Hunt, Jmarine and P.T. Stone. Decision trees are classifiers which consist of a set of rules which are applied in a sequential way and finally yield a decision. They can be best explained by observing the training process, which starts with a comprehensive training set. It uses a divide and conquer strategy [35][40].

6. Sentiment Classification Using Unsupervised Learning

Since sentiment words are often the dominating factor for sentiment classification, it is not hard to imagine that sentiment words and phrases may be used for sentiment classification in an unsupervised manner. It performs classification based on some fixed syntactic patterns that are likely to be used to express opinions. The syntactic patterns are composed based on part-of-speech (POS) tags...

Cross-language sentiment classification means to perform sentiment Classification of opinion documents in multiple languages. However, much of the research has been done in English, while there are not many resources or tools in other languages that can be used to build good sentiment classifiers quickly in these languages. The second motivation is that in many applications, companies want to know and compare consumer opinions about their products and services in different countries. If they have a sentiment analysis system in English, they want to quickly build sentiment analysis systems in other languages through translation. In (Wan, 2008), the author exploited sentiment resources in English to perform classification of Chinese Reviews. Wei and Pal (2010) proposed to use a transfer learning method for cross language sentiment classification.

IX. COMPARATIVE ANALYSIS OF MACHINE LEARNING TECHNIQUES

The techniques used are language dependent and thus cannot be applied to sentiments in other languages. Contradictory to this, a language independent Sentiment Analyzer can analyze Sentiments expressed in many languages. These techniques do not use language specific features in classifying the subjectivity in the sentiment. Thus, they are generalized and can be applied to documents in different languages. Another important aspect of Sentiment Analysis is whether the technique uses a training set. Training Set is a tagged dataset. Already tagged sentiments are also sometimes used as training set. Training set helps in the construction of the model. Sometimes sentiment associated with words may be different in different domains [44][45]. There is not even a single technique which provides efficient result for every product. Different supervised categorization algorithms have been used so far for polarity classification tasks. Most commonly used methods are Support Vector Machine (SVM), Naïve Bayesian Classifier. While other methods like Maximum Entropy, Decision Tree, also used for the same task. We studied many sentiment analysis algorithms, but we find that the Support vector machine is the most reliable and efficient algorithm. SVM provides better accuracy than any other algorithm. There are four values that are commonly used to evaluate performance of text categorization and opinion mining in machine learning approach, namely accuracy, recall, precision and F measure [46]. SVM is used in most of the sentimental approach because it provides the much accurate result than any other machine learning technique. We studied more than 300 algorithm from 2002 to 2013 we find that most of the algorithms used support vector machine learning technique.

<table>
<thead>
<tr>
<th>S.N.</th>
<th>Technique</th>
<th>Accuracy %</th>
<th>Use %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SVM</td>
<td>75-85</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>NVB</td>
<td>65-75</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>M aximum Entropy</td>
<td>60-65</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>Decision Tree</td>
<td>40-60</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
<td>Below 40</td>
<td>5</td>
</tr>
</tbody>
</table>

Table -1

SVM with a lower classification provide accuracy up to 86.70%, while 87.87% for positive values and 85.53 for negative value [48]. Samsudin, puteh, hamdan &nazri(2012) proposed a hybrid algorithm artificial immune system which gives better accuracy than the SVM for movie review[51].
X. SENTIMENT ANALYSIS TOOLS

There are various sentiment Analysis tools available in the market every tool has its own importance and developed for a specific product. Following are the sentiment analysis tools:

1. Earshot: Get real-time sentiment data filtered by location.
2. VendAsta: See how sentiment trends have shifted over time.
3. Radian6: Use sentiment analysis to streamline social media workflows.
4. Swipp: Gauge reaction to digital marketing campaigns.
5. Yext: Get notified when sentiment starts to fall.
6. ListenLogic: Clean up your sentiment data [49]
7. People Browser: Find all the mentions of your brand, industry and competitors and analyze sentiment.
8. T weetstats: T his is a fun, free tool that allows you to graph your T witter stats. Simply enter your T witter handle and “let the magic happen.”
9. Hootsuite: A great freemium tool that allows you to manage and measure your social networks.
10. Marketing Grader: It uses over 35 metrics to calculate your grade by looking at if you are regularly blog posting[50]

XI. SENTIMENT ANALYSIS APPLICATION

Sentiment analysis has its application importance in each and every industry for the customer and company prospective. We have already discussed its importance as “A need of an organization and Requirement of a customer” so Sentiment analysis is useful for every organization and user. Following are the major application of sentiment analysis and opinion mining.

1. Applications to Review-Related Websites
2. Applications as a Sub-Component Technology
3. Applications in Business and Government Intelligence
4. Applications Across Different Domains
5. The analysis of patents developed to a large application area
6. Text Classification for News Agencies
7. Bio-informatics-Bio-entity recognition aims to identify and classify technical terms in the domain of molecular biology that correspond to instances of concepts that are of interest to biologists
8. Anti-Spam Filtering of Emails Opinion Poll decision making [6][35].

XII. CHALLENGES

Although sentiment analysis is the most interesting field of researcher’s, since 2000 many problems have been solved by researchers, but still machine learning is not as efficient as a human being. Following are the major issue and challenge in sentiment analysis and opinion mining. Although sentiment words and phrases are important for sentiment analysis, only using them is far from sufficient. The problem is much more complex. In other words, we can say that sentiment lexicon is necessary but not sufficient for sentiment analysis.

Following are the main Challenge in sentiment analysis.

1. A positive or negative sentiment words may have opposite orientations in different application domains.
2. A sentence containing sentiment words may not express any sentiment Example interrogative sentences and conditional sentences.
3. Sarcastic sentences with or without sentimental words are hard to deal with, e.g., “What a great car! It stopped working in two days.”
4. M any sentences without sentiment words can also imply opinions. M any of these sentences are actually objective sentences that are used to express some factual information example- “Wow! What a good camera its zoom quality is so bad”. [6][33].
5. Opinion Spams, many organization post spam about their products which also a big issue in analyzing sentences.
6. Object Identification, feature Extraction, Synonym grouping and integration is also an issue in sentiment analysis [33].
7. Hidden Sentiment Identification (like anger, disgust and joy is a challenging task) [44].

XIII. CONCLUSION

Sentiment analysis and opinion mining is an emerging research field of Data Mining and Natural language processing. We introduce a brief literature survey and, existing research work and major issue in order to find the clear direction to the future work. We explain how the sentiment analysis is useful in Market research for the customer and company prospective. We also surveyed the existing platform and computational approaches that are being used in the current scenarios. To complete this survey we gone through more than 200 papers and collect all the useful information so the follower of this paper will understand the proper way to study the sentiment analysis and opinion mining.

REFERENCE
