IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

A Study and Survey on Sentiment Analysis

Kajaree Das

B. Tech Student, Department of Information Technology, Institute of Engg and Management, Kolkata, India

Abstract: In essence, Sentiment Analysis (SA) is detection and determination of the response of the targeted consumers to a certain brand or product or maybe even a situation. But there are much more potential in SA and emerging research in this area has attracted and involved many brilliant academic minds till date. Human mind is biased with preferences and judgements. So, automated machine to identify and clarify opinions presented in electronic unstructured text has come into picture and became the main focus of present research. There are also some challenges present in this field, such as, accuracy. Further ongoing researches are focusing on solving these problems and creating more efficient tool for SA.This paper tried to cover the study and research on SA that has been done so far. It also talks about the challenges and future aspects of SA.

Keywords: Sentiment, Analysis, Opinion, Data, Lexicons, Level.

I. INTRODUCTION

Man is a social animal. Opinions and sentiments of others have always been important and crucial to him and his decision making. So, Sentimental Analysis shares an ancient history with the society.
But the time has changed and the civilisation has evolved. So has the society and everything and everyone related to it. We live in a digitalised world, to which internet is a lifeline. We live and breathe on social media. The virtual communities have grown into a valuable place where different people shares their ideas, thoughts and opinions on various topics, products, brands, different socio-economic subjects and what not. So practically those virtual communities mirror the real life scenario in essence. And wherever views and opinions are concerned Sentiments are involved. Hence, Sentimental Analysis has become an integral part of social listening and an emerging research interest.

Different Scientific communities (for challenging research topics) and the business world (for market prediction) are showing a blooming interest in gathering public opinion about society, politics, marketing campaigns of different brands and products etc. So Opinion Mining and Sentimental Analysis have emerged and various automated tools and algorithms have been developed.

A. The Beginning

Although this area of research has gained popularity recently but it would be impossible to deny the undercurrent of interest that was growing from early 90's. In 1979, Jaime Guillermo Carbonell published his Ph.D. thesis on Subjective Understanding: Computer Models of Belief Systems [32]. There is another paper worth mentioning which was published in 1984.[] Then there was an array of papers published which focused on mainly metaphor, narrative, perceptions, affection, encoded texts and other related areas.
But the major research on the challenging problems and probable applications of Sentiment Analysis and Opinion Mining did not surface until 2001. Factors that enhanced this sudden burst of research interest can be summed up as:

- The introduction of machine learning approaches in NLP and information extraction
- Availability of a vast amount of data due to development and rise of social media, e-commerce sites, review websites, blogs etc.
- Widespread opportunities and applications these area potentially offers.

A. Sentimental Analysis and / or Opinion Mining

Though these two terms are commonly used as synonyms and both of them use NLP algorithms to detect, extract and distil opinions from various social media, blogs, review sites etc., they represent slightly different field of study. Opinion Mining focuses on detecting the polarity of opinion (positive or negative) and Sentimental Analysis tried recognise the emotions. But as recognising sentiments are often used to detect polarity of a statement, these two fields are combined together and commonly used to refer to same field of study.

B. Main Fields of Research

The major researches in SA can be categorised into the following sub-fields:

1. Subjectivity Detection: Determining the text is opinionated or not.
2. Sentiment Prediction: Predicting the polarity (positive or negative) of the text
3. Aspect Based Sentiment Summarisation: Summarisation of the sentiments in quantitative form, i.e., ratings or scores
4. Text Summarisation: Extracting the essence of a long review in very few lines.
5. Contractive Viewpoint Summarisation: Deals with contradictory opinions about a particular topic.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

6. Product Feature Extraction: Distillation of the product feature from its reviews.
7. Opinion Spam Detection: Recognising the fake, biased and judgemental opinions from reviews.

D. **CHALLENGES**:
Irrespective of all the benefits of SA, there are some challenges faced by all the researchers. Firstly, polarity of an opinion word may vary with the situation concerned. Secondly, most of the people express their sentiment in their own ways which may significantly differ from others. So, formatting an automated tool to detect the emotions is not simple or easy.

Some statements can be contradictory or contain mixed emotion making it very difficult to parse the information by computer even though the statement was easily understandable for any human being to begin with. Some statements are even not clearly understood by other people for lack of context. Also, some opinions are biased as sometimes people like to go with their own preferences rather than depending on the facts or logic.

The researchers also face certain challenges same as those in NLP, such as, co-reference resolution, negation handling, anaphora resolution, named-entity recognition and word-sense disambiguation. The challenges and problems in SA are a major portion in the related research and so we have discussed them elaborately in Chapter-VII.

E. Languages
The Natural languages those have been mostly studied is English and Chinese. At present there are very few ongoing researches on languages like Arabic, Italian and Thai. This survey paper aims to focus on research conducted on English and Chinese. There are also some ongoing researches on Indian languages which are briefly listed in Chapter - VIII.

II. SOURCE OF DATA: THE WORLD WIDE WEB

The opinions of targeted consumers is very important for any level of business. In World Wide Web, personal blogs, community blogs, social media, virtual communities and micro-blogs constitutes the data mine proper analysis of which provides a good understanding of not only the present but also the future market status of a product or service or a brand.

A. Blogs:
Blogs are one of the most popular and major medium through which individuals share their personal opinions with the world. Some bloggers keep daily record of their life; share their own thoughts, emotions and views with the readers. Many bloggers focuses on reviewing certain brands, newly launched products, social issues etc. These blogs are considered to be one of the valuable resource and many studies related to SA have used blogs to collect unstructured opinionated text.

B. Review Sites
Many buyers consult the reviews present over the internet before buying any new product or trying a new recipe at home or even visiting a new restaurant. At resent most of the data used in SA are being collected from e-commerce websites like Amazon and Flipkart (product reviews), Zomato (restaurant reviews), Tripadvisor (Travel guide and hotel reviews) and many other popular sites which include millions of product or service reviews and ratings by the consumers.

C. Data Set
There are certain sites available which collect raw data from the blogs or review sites and keep the data online for researchers to access those data without much hassle. Like, Movie review data are available on www.cs.cornell.edu/People/pabo/movie-review-data. Multi Domain Sentiment Dataset (MDS, www.cs.jhu.edu/mdredze/datasets/sentiment) includes four different type product (Books, DVDs, Electronics and Kitchen Appliances) reviews extracted from amazon.com.

D. Micro-blogging:
The most popular micro-blogging site is Twitter where users share their views in short messages known as "Tweets". These tweets, being opinionated texts, play a major role in sentiment classification.

III. DIFFERENT LEVELS OF SENTIMENT ANALYSIS

There are three key levels of approach in SA:

A. Document Level
Analysing the overall emotion conveyed about any topic (target) in the text. It assumes the whole comment or statement discusses only one topic. Hence, evidently it fails to investigate the sentiment of the document if multiple entities are involved. This technique cannot be used in case of blogs and forums where comparison of different yet similar product may be present in the text. Also, at this level of sentiment classification different sentiments about different aspect of the entity cannot be separated. For Document Level SA, both supervised learning algorithms (e.g., naïve Bayesian, Support Vector Machine) and unsupervised algorithms (Collecting the opinions or sentiments in a single document) can be employed.

B. Sentence Level
Parsing the sentiment expressed in each sentences those are present in the comment or text. This level of analysis is close enough to subjectivity classification [], which separates objective sentences or sentences containing factual information from subjective sentences expressing subjective views or opinions. Researchers also opted for clause level and phrase level analysis. But that did not fruit well either because determining the polarity of a sentence is of lesser use than knowing the polarity of opinion about a particular feature or aspect of a product or topic.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

C. Entity and Aspect Level

It employs more finer-grained approach than the others. It analyses by considering each sentiment or opinion present in the content. Researchers used to call this feature level analysis earlier. [] This level of analysis is based on the idea that an opinion roughly is an expression having two main components:

A target (commonly referred to as 'topic' or 'entity' by most social Analytics tools)
A sentiment is determined on the topic.
So, in "I love this shampoo", "this shampoo" is the topic, and the sentiment or emotion (as expressed by "love") is positive.

Entity and Aspect Level analysis is really helpful because of the fact that it can summarise an unstructured text or data into a structured description about sentiments about a target. But both document level and sentence Level SA is difficult and complicated to implement, and entity and aspect level analysis makes it even more challenging. And when we started to separate regular opinions from comparative opinions the complications increases even further.

IV. FEATURES

To classify and analyse sentiments, the necessary first step is to identify the sentiments, classify the opinion words according to their features (positive, negative or neutral) and extract them from the available unstructured electronic text. We currently look for the following features in the targeted text:

A. Terms and Frequency

The most common features used in traditional sentiment classification are words (unigram) and their N-grams with observed frequency. In some cases positions of words may also be accounted for. The TF-IDF weighing method are also employed sometimes. This either verify the words appear or not or uses the frequency count for the associated terms to compare the importance of the features.

B. Parts of speech

Parts of Speech is another important feature which is used by many researchers to clarify and distil the sentiment associated with the word. They have considered adjectives to be a special feature as adjectives are major indicators of opinion. However, theoretically one can use the POS-tags of all the word and their N-grams to classify the features.

C. Sentiment Words and Phrases

These are words or phrases which are used to express sentiments or opinions be it positive or negative. Most of the opinion or sentiment words are adjective (wonderful, terrible etc.) and adverbs. Few of them are noun (rubbish, junk etc.) and verbs (love, hate etc.). There are some common opinionated phrases which do not include any opinion words, like, "cost me an arm and leg." There are also sentiment idioms.

$$\chi_i^2 = \frac{n \cdot F(w)^2 \cdot (p_i(w) - P_i)^2}{F(w) \cdot (1 - F(w)) \cdot P_i \cdot (1 - P_i)}$$

D. Sentiment shifters

There are certain words which when used change or shift the polarity of the opinion. For example, 'I like this place.' Is positive but we can make it negative ('I don't like this place.') adding a 'don't'. So, Negation words are one of the most common example Sentiment Shifters. This shifters need to be dealt with care because their every occurrence does not necessarily mean changing or shifting in sentiment orientation. Like, sentences having the clause 'not only…..but also' express usually positive emotion even a negation word 'not' is present.

E. Syntactic Dependencies

The researchers also focused on the dependency-based features of words or dependency tree generated while parsing a text.

V. FEATURE SELECTION METHODS

There are two main feature selection methods: Lexicon-based and Statistical methods.

A. Lexicon-based Methods

This method needs human annotation. The first step of this approach is selection of a small set of 'seed' words. Then this set is used to synonym detection to formulate a larger lexicon. This approach has many difficulties [] and is not frequently used.

B. Statistical Methods:

These methods are most commonly used and fully automatic. This approach classify the whole document into several groups of words known as Bag of Words (BOW) or strings (sequence of words). BOW is much simpler and used more frequently. Some commonly used statistical FS methods are:

1. Point-wise Mutual Information

The basic Point-wise Mutual Information (denoted by $M_i(w)$) is a formal measure which denotes the degree of co-occurrence between a class i and a word w.
$M_i(w) = \log(F(w) \cdot p_i(w) / F(w) \cdot P_i(w)) = \log(p_i(w) / P_i(w))$
If $M_i(w) > 0$, the correlation is positive and otherwise its negative.

Yu and Wu have developed a contextual entropy model by enhancing the basic PMI model. They considered contextual distribution of words together with the co-occurrence strength to acquire emotion words from stock market news articles. This model compared the contextual distribution of words to identify the word or words of the same kin as the seed words. Then the seed words and those akin words were together used to classify the targeted document. This method was proven to be more efficient in extracting more useful opinion words and less noisy words than the basic PMI method.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

2. Chi-square

χ^2 is another way of measuring the correlation between terms or words and categories or classes. Suppose, we have n documents to consider. P_i is the probability that document contain the class i and F(w) is the fraction of documents those have the word w. $p_i(w)$ denotes the probability that class I contain the word w. Then,

Fan and Chang presented one application of the chi-square statistic in finding correlation between advertisements and the personal interests of bloggers.

3. Latent Semantic Indexing

This is one of the most popular method for drawing a smaller sample of features from the actual set there by transforming the text space into a new vector-space system which combines all the original features linearly. This method does not analyse the meaning of the text but identifies the patterns those are present in the documents.

V. SENTIMENT CLASSIFICATION A LITERATURE SURVEY

After going through all the research work on Sentiment analysis done over the past few years, it is inferred that SAinvolves several overlapping stages. But they can broadly be identified into five main tasks [11]:

A. Process of Sentiment Analysis for Text (Lexicon Generation)

In the first stage, target words are filtered out and lexicons are generated to closely analyse the sentiments or opinions attached. According to ongoing researches, prior polarity should be attached at each lexicon level. To develop SentiWordNet(s) in different languages, both automated and manual processes have been applied. In the year 1966, Philip Stones designed an automated system named General Inquirer and set the first milestone for automatic extraction of textual sentiment. The system used databases containing set of words those were manually classified as sentimentally positive or negative. The words are referenced with this database to classify their sentiment orientations such as positive, negative, feel, pleasure [5]. In the year 1994, Brill Tagger proposed the semantic orientation based of parts of speech. In the year 1997, Hatzivassiloglou was the first to propose an empirical process of generating sentiment lexicon for adjectives. The algorithm made use of the connectors joining the adjectives. A logarithmic linear model of regression that provided 82% of accuracy [8]. But for better and faster analysis of opinions, there was growing need of a fully automated system, which could be useful for any kind of electronic documents. Later in the year 2002, Turney developed an Algorithm to identify semantic or sentimental polarity of target phrases [31][7]. He proposed an idea of Thumbs up (positive opinion) and Thumbs down (negative opinion). For extracting the resulting polarity of consecutive words, Turney designed an algorithm to extract Point wise Mutual Information (PMI). In 2002 itself, Pang developed sentiment lexicons for movie review data corpus to indicate the polarity of opinion. This system inspired the supervised machine learning techniques like Naive Bayes, Maximum Entropy and Support Vector Machine (SVM)[10].

In 2004, Kamps, Marx, Mikken and Rijke tried to determine the influence and contextual dependencies of adjectives in Word Net. In their research, the adjectives were categorised into four basic classes and base words were used (to calculate relative distance) depending on the class. For example, class Feeling had base words “happy” and “sad”, class Competition had base words “pass” and “fail”, etc. Based on this concept, they managed to classify a total of 1608 English words into four classes and the average accuracy of classification was 67.18% [24].

Gamon in 2005 proposed a method similar as what Turney proposed in 2002. This algorithm was based on Machine Learning approaches and used with input of seed words. This classifier assumed that the target words with similar polarity might co-occur in one sentence but words with different or opposite polarity could not [11].

In 2005, Read came up with another three different issues in the area of sentiment analysis. According to him Time, Domain and dependency of words on the topic vastly effect sentiment orientations. He showed that associative polarity of sentimentmight change with time [12].

In 2009, Denecke introduced SentiWordNet a resource of data corpus already assigned prior polarity scores. He proposed two approaches: rule-oriented and machine learning oriented. Accuracy of the rule-based algorithm (74%) was less than that of machine learning based (82%) [13].

Mohammad in 2009, proposed a new technique to further increase the extent of sentiment lexicon. It included the recognition of individual words as well as phrasal expressions with the help of a thesaurus and a list of affixes. The implementation of this technique involved two approaches: Thesaurus based and antonymy generation based. Manually designed rules were used for antonymy generation based method. Thesaurus method involved a seed word list.According to this method,if a document or a paragraph contained more positive seed words than negative, then paragraph is marked as positive [14].

Mohammad and Turney in collaboration designeda web service to obtain human annotation of opinion lexicon named Amazon Mechanical Turk. Various validations followed as to eliminate or re-annotate erroneous and out of context annotations [10].

B. Subjectivity Detection

Subjectivity of the targeted text refers to the underlined opinion of the same where as the objectivity refers to the fact content. More precisely,subjectivity could be termed as the Topical Relevant Opinionated Sentiment [9]. Genetic Algorithm proposed by Das in 2011 achieved a significant success to detect the subjectivity for Multiple Objective Optimisation [27].

An example-

1. Subjective - The movie ‘A Beautiful Mind’ was really enjoyable. (Opinionated sentence; hence subjective)

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

2. Objective- 'A Beautiful Mind' depicts the life of the famous scientist John Nash. (Fact stating sentence; hence objective).
Wiebe in 2000, explained the concept of subjectivity from the perspective of information retrieval which involved the two categories subjective and objective [9].
In their paper published in 2005, Aue and Gamon described the problem of subjectivity identification dependent on both context and domain dependent. This revelation directly contradicts with the use of conventional databases with words assigned with prior polarity scores like SentiWordnet or subjectivity word list etc. [17].
Das and Bandyopadhyay (2009) further explained the approaches for determining subjectivity using Rule-based approaches, Machine learning techniques and Hybrid phenomenon [2]. The idea of using a collection of subjectivity clues in turn helped to detect subjectivity in target phrases and sentences. This collection included adjectives, verbsand n-grams [8][9][18].
Zhao gave the idea of using machine learning algorithms like Support Vector Machine (SVM), Conditional Random Field(CRF) to form clusters of similar types opinions[6].

C. Sentiment Polarity Detection
The detection of polarity of sentiments involves classifying the words into semantic classes (Turney et.al.,2002) (e.g. positive, negative or neutral) and other emotional annotations like angry, sad, happy, surprised [7].
For the last several years, some online sentiment analysis tool like Tweet Feel (www.tweetfeel.com) Twitter Sentiment Analysis Tool (twittersentiment.appspot.com/) are available. But the level of research involved is not satisfactory [19].
Cambria in 2011 designed a new techniquecalled Sentic Computing. It is a multi disciplinary approach involving on common sense reasoning and emotion extraction. It had used short texts or messages to interpret social information available over the web [20].
Concept Net is another semantic network initially introduced with approximately 10000 concepts and 72000 or more features taken from Open mind corpus.

D. Sentiment Structurization
So far, the explanation of Sentiment Analysis did not consider the end user although the main purpose and applications of sentiment analysis solely depend on the need of end user. The needs of the end users are not limited in identifying only positive or negative opinions but how that underlying sentiment is significant to him or her. For example, an e-commerce site may be interested to track its buyers' preferences or feedbacks for future marketing policies. So, the site may employ certain sentiment analysis algorithms that produce an aspectual output for its reference.

To address this problem, an approach known as sentiment structurization was proposed by Das (2010). This method involves answering 5Ws (Why, Where, When, What, Who).The main drawback of 5Ws was that it might induce some biased label. To minimise the probability of bias induction, Maximum Entropy Model (MEMM) was proposed.

Bethard first introduced the concept of automatic determination of opinions by answering questions in 2006. Next year, Bloom described the Appraisal Theory (Martin and white, 2005).The system categorises the sentiments into three classes: affect, appreciation and judgment. Yi introduced an online sentimentanalyzer for text documents available over the world wide web. Zhou designed the structure for blogosphere to summarise the text.

E. Sentiment Summarization-Visualization-Tracking
One of the most importantstep to meet the needs of users is the interpreting of data. Two types of summarisation attempt are introduced:
□ □ Polaritywise (Hu, 2004), (Yi and Niblack, 2005), (Das and Chen, 2007)
□ □ Topicwise (Yi et al., 2003), (Pang and Lee, 2004), (Zhou, 2006)

Visualisation and Tracking are the final phase of sentiment analysis. Butits of utmost important for this phase produce the result that is shared with the end user. This phase involves generation of visual sentiments which are represented graphically polarity wise according to some particular dimension(s).

VII. CHALLENGES

The implementation of Sentiment Analysis is not devoid of challenges. Over the past few years researchers faced and devoted their work to solve those challenges. Some of those are discussed below.

A. Named Entity Identification
Named entities can be defined as nouns or phrases that refer to any particular entity like an organisation, a person, a commodity, and so on. The main purpose of identifying such entity is resolve all of its textual mentions in a text piece to interpret the underlying sentiment about that entity depicted by the piece. Classifier based Sentiment Analysis is well suited for this task. But it always not easily identifiable that about which entity the text is precisely pointing at. For example: 'Is Matrix a mathematical concept or movie?'

B. Review Structure
The construct of the available text effects sentiment analysis deeply:
Structured data or texts are found in formal reviews written by professional writers about books, scientific issues and so on.
Semi-Structured texts can be put somewhere intermediary between the formally structured reviews and unstructured reviews. For example a review in the form of Pros and Cons where those are written in short phrases rather than complete sentences.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

Unstructured data or reviews referred to reviews written in an informal format. Like: a Facebook post or Tweets with emoticons, abbreviations, grammatically incorrect sentences etc.
A compact tool that can be implemented for all these types of text format is yet to be available.

C. Resolution of Anaphoric References
To identify what exactly a pronoun or a phrase referring to in a sentence involving multiple entities is another major challenge in sentiment analysis. For Example: 'The children went to the zoo and then the museum; it was awesome.' The question arises what 'it' refers to.

D. Presence of Sarcasm
Often people use sarcastic sentences to prove their point or lay down an argument. In those cases the assignment of polarity to a particular text becomes more difficult as the conventional polarity assigned data corpus failed to reach the actual conclusion.

E. Relation Between Words
The semantics of a sentence can be determined if the relationship and dependencies between the words are identified. A majority of researchers has focused on solving this problem in NLP. But due to presence of unstructured text elements, the analysis is difficult and available methods do not always produce error free result.

F. Domain Dependency
Machine learning algorithms are used to design sentiment classifiers and the sentiment classifiers are trained using a specific set of data. A classifier thus trained can produce highly inefficient result if employed to a different domain of data. For example, a classifier designed for classifying sentiments about a movie may fail to classify sentiments about cars to a certain satisfaction level. Thus domain dependent algorithms are preferable to give better result than domain independent classifiers.

VIII. SENTIMENTAL ANALYSIS ON INDIAN LANGUAGES

One of the most difficult challenge in implementation of sentiment analysis may be the language diversity all over the world. The main researches focused on more popular languages like English, Spanish etc. So the researches that have been conducted Indian languages are far less as expected but their significance and relevance cannot be ignored none the less.

In 2002, Narayan used Hindi WordNet anda Subjective Lexicon in Hindi to identifythe semantic orientation of adverbs and adjectives [25].
In 2004, Kim and Hovy worked on sentiment analysis for Hindi Language but the work was strictly limited to synonyms only[24].
In the year 2010, Das and Bandyopadhyay took the first major initiative to crate a SentiWordNet in Bengali using a bilingual dictionary (English-Bengali) and Sentiment Lexicons that were available in English[21].
They proposed a four staged approach to assign polarity to a word.

1. interactive game todefine the polarity of words
2. bi-lingual dictionary for English and a particular Indian Languages
3. word net is developed using antonym and synonym relations
4. a previously annotated data corpus is used for unsupervised learning [1]

They designed the technique for word tagging using Bengali words. Words were broadly classified into six categories namely happy, sad, surprise, fear, disgust and anger with intensities among low, general and high [22].
Beside that, in the same year, Joshi came up with Hindi - SentiWordNet (H - SWN) using lexical resources English SentiWordNet and English- HindiWordNet Linking.

IX. CONCLUSION AND FUTURE SCOPES

This paper describes the basic concept of Sentiment Analysis as well as gives an overview of most of the relevant researches in this field. The main challenge in Sentiment Analysis still remains to be the unstructured nature of data or texts available.
In spite of a rich history and plenty of researches, many areas of sentiment analysis are yet to be explored. Also the linkage between research and application is still weak. A complete sentiment classifying tool with significant accuracy is yet to be designed. So the future scopes are endless. Also, the interest in implementing Sentiment Analysis in natural languages other than popular ones (e.g. English, Chinese etc) is growing but the ongoing research works and available resources are still limited. In most of the situation, the context of the text and preferences of end user are to be considered. So the future researches may focus on addressing context based and cross - domain Sentiment Analysis.
As for this study and survey, a complete and satisfactory study of all the machine learning algorithms available, both simple and hybrid, is yet to be published. Also there are several ongoing researches on implementing Nature Inspired Algorithms to solve Sentiment Analysis. A survey of that will also be helpful for future references.

REFERENCES

[1] A. Das and S. Bandyopadhay, "SentiWordNet for Indian languages," Asian Federation for Natural Language Processing, China, pp. 56–63, August 2010.
[2] A. Das and S. Bandyopadhay, "Subjectivity Detection in English and Bengali: A CRF-based Approach," In Proceedings of the 7th International Conference on Natural Language Processing, Macmillan 2009.
[3] H. Tang, S. Tan and X. Cheng, " A survey on sentiment detection of reviews", In Proceedings of the Expert Systems with Applications 36, Elsevier Ltd., Beijing, 2009
[4] N. Mohandas, J.P. Nair and G.V, " Domain Specific Sentence Level Mood Extraction from Malayalam Text", In Proceedings of

the International Conference on advances in Computing and Communications, IEEE, 2012.
[5] P.J. Stone, "The General Inquirer: A Computer Approach to Content Analysis", The MIT Press, 1966.
[6] J. Zhao, K. Liu and G. Wang, "Adding Redundant Features for CRFs- based Sentence Sentiment Classification" In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp.117-126, 2008.
[7] P. Turney, "Thumbs up or thumbs down? Semantic orientation Applied to Unsupervised Classification of Reviews" In Proceedings of the Association for Computational Linguistics, pp.417-424, Philadelphia, 2002.
[8] V. Hatzivassiloglou and K. R. McKeown, "Predicting the Semantic Orientation of Adjectives", In Proceedings of the 35th Annual Meeting of the ACL and 8th Conference of the European Chapter of the ACL, pp.174-181, Madrid, 1997.
[9] J. M. Wiebe, " Learning Subjective Adjectives from Corpora", In Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence, pp. 735- 740, Mento Park, 2000.
[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques" In Proceedings of the Empirical Methods on Natural Language Processing, pp. 79-86, Pennsylvania, 2002
[11] M. Gamon, A. Aue, S. Corston-Oliver and E. Ringger, "Pulse: Miming Customer Opinions from Free Text", In Proceedings of the International Symposium on Intelligent Data Analysis, pp. 121-132, 2005.
[12] J. Read, "Using Emotions to Reduce Dependency in Machine Learning Techniques for Sentiment Classification", In Proceedings of the Student Research Workshop, pp. 43-48, Arbor, 2005.
[13] K. Denecke, "Are SentiWordNet Scores Suited For Multi-Domain Sentiment Classification", In Proceedings of the 4th International Conference on Digital Information Management, pp. 33-38, Ann Arbor, 2009.
[14] S. Mohammad, B. Dorr, and C. Dunne, " Generating High- Coverage Semantic Orientation Lexicons fom Overly Marked Words and a Thesaurus", In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pp. 599-608, 2009.
[15] S. Mohammad and P. Turney, "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon" In Proceedings of the Workshop on Computational Approaches to Analysis and Generation of Emotion, pp. 26-34, California, 2010.
[16] J. M. Wiebe, "Recognizing Subjective Sentences: A Computational Investigation of Narrative Text", Doctoral Thesis. UMI Order Number: UMI Order No. GAX90- 22203., State University of New York, 1990.
[17] M. Gamon and A. Aue, " Automatic Identification of Sentiment Vocabulary: Exploiting Low Association with Known Sentiment Terms", In Proceedings of the Workshop on Feature Engineering for Machine Learning in Natural Language Processing, pp. 57-64, Ann Arbor, 2005.
[18] K. Dave, S. Lawrence and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews", In Proceedings of 12th International Conference on World Wide Web, pp. 519- 528, Hungary, 2003.
[19] B. Liu, "Sentiment Analysis: A Multi- Faceted Problem", In Proceeding of IEEE Intelligent Systems, pp. 76-80, 2010. E. Cambria, A. Hussain and C. Eckl, " Taking Refuge in Your Personal Sentic Corner", In Proceeding of Workshop on Sentiment Analysis where AI meets Psychology, pp. 35-43, Thailand, 2011.
[20] A. Das and S. Bandyopadhay, "SentiWordNet for Bangla", 2010. A. Das and S. Bandyopadhay, "Labeling emotion in Bengali blog corpus a fine grained tagging at sentence level", In Proceeding of 8th workshop on Asian Language Resources, pp. 47-55, Beijing, 2010.
[21] G. Draya, M. Planti, A. Harb, P. Poncelet, M. Roche and F. Trousset, " Opinion Mining from Blogs", In Proceeding of International Journel of Computer Information Systems and Industrial Management Applications, 2009
[22] S. M. Kim and E. Hovy, "Determining the sentiment of opinions", In Proceeding of COLING, pp. 1367-1373, 2004.
[23] D. Narayan, D. Chakrabarti, P. Pande and P. Bhattacharyya, " An experience in building the indo wordnet -a wordnet for hindi", In Proceeding of First International Conference on Global WordNet, 2002.
[24] D. Rao and D. Ravichandan, " Semi-supervised polarity lexicon induction", In Proceeding of 12 conference of the European Chapter of the Association for Computational Linguistics, pp. 675-682, USA, 2009.
[25] A. Das, "Opinion Extraction and Summarization from Text Documents in Bengali", Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., Jadavpur University, 2011.
[26] A. Joshi, A. R. Balamurali and P. Bhattacharyya, "A Fall- back Strategy for Sentiment Analysis in Hindi: a Case Study", In Proceedings of the 8th ICON, 2010.
[27] Amitava Das, http://www.amitavadas.com/
[28] Janyce M. Wiebe,"Learning Subjective Adjectives from Corpora", http://www.cs.columbia.edu/~vh/courses/LexicalSemantics/Orientation/ wiebe-aaai2000.pdf
[29] Brill Tagger, http://www.ling.gu.se/~lager/mogul/brill-tagger/index.html.
[30] Jaime GuillermoCarbonell, "Subjective Understanding: Computer Models of Belief Systems", Ph.D. thesis, Yale University, USA
[31] Aggarwal Charu C, Zhai Cheng Xiang. Mining Text Data. Springer New York Dordrecht Heidelberg London: Ó Springer Science+Business Media, LLC'12; 2012.
[32] Teng-Kai Fan and Chia-Hui Chang, "Sentiment-Oriented Contextual Advertising," Knowledge and Information Systems, vol. 23, no. 3, pp. 321-344, 2010.
[33] Raisa Varghese and Jayasree M, A Survey On Sentiment Analysis and Opinion Mining, IJRET, Vol. - 2, Issue 11, Nov, 2013
[34] Amandeep Kaur and Vishal Gupta, A Survey On Sentiment Analysis and Opinion Mining Techniques, Journal Of Emerging Technologies in Web Intelligence, Vol. 5, 2013
[35] G. Vinodhini and RM. Chandrasekaran, Sentiment Analysis and Opinion Mining: A Survey, IJARCSSE, Vol 2, Issue 6, June 2012
[36] Walaa Medhat, Ahmed Hassan and Hoda Korshy, Sentiment Analysis Algorithms and applications: A Survey, Ain Shams Engg. Journal, 2014
[37] Hussein, D.-M.E.D.M. A survey on sentiment analysis challenges. Journal of King Saud University – Engineering Sciences (2016)
[38] Teng-Kai Fan, Chia-Hui Chang , "Blogger-Centric Contextual Advertising" Expert Systems with Applications 38 (2011) 1777–1788 .Ting-Chun Peng and Chia-Chun Shih , "An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and intelligent Agent Technology JOURNAL OF COMPUTING, VOLUME 2, ISSUE 8, AUGUST 2010, ISSN 2151-9617 .
[39] Tsur, D. Davidov, and A. Rappoport , " A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews". In Proceeding of ICWSM. of Context Dependent Opinions,2010.
[40] Weitong Huang, Yu Zhao, Shiqiang Yang, Yuchang Lu, "Analysis of the user behavior and opinion classification based on the BBS", Applied Mathematics and Computation 205 (2008) 668–676 .
[41] Wiebe, W.-H. Lin, T. Wilson and A. Hauptmann, "Which side are you on? Identifying perspectives at the document and sentence levels," in Proceedings of the Conference on Natural Language Learning (CoNLL), 2006.
[42] Yi and Niblack, "Sentiment Mining in Web Fountain",Proceedings of 21st international Conference on Data Engineering, pp. 1073-1083, Washington DC,2005.
[43] Yongyong Zhail, Yanxiang Chenl, Xuegang Hu, "Extracting Opinion Features in Sentiment Patterns", International Conference on Information, Networking and Automation (ICINA), 2010.
[44] YuanbinWu, Qi Zhang, Xuanjing Huang, LideWu, "Phrase Dependency Parsing for Sentiment analysis", Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 1533–1541, Singapore, 6-7 August 2009 .
[45] ZHU Jian, XU Chen, WANG Han-shi, "" Sentiment classification using the theory of ANNs", The Journal of China Universities of Posts and Telecommunications, July 2010, 17(Suppl.): 58–62.