National Conference on Engineering Innovations and Solutions (NCEIS – 2018)
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 4 | Issue 6 | ISSN : 2456-3307

A Survey on Sentiment Analysis and It's Applications

Harshitha G[1], Lavanya G[1], Hemanth Kumar A[2]
[1]Department of CSE, G. Madegowda Institute Of Technology (GMIT), Maddur, Mandya District, Karnataka, India
[2]Assisstant Professor, G. Madegowda Institute Of Technology (GMIT), Maddur, Mandya District, Karnataka, India

ABSTRACT

Sentiment analysis refers to one of the particular task which is included under natural language processing and it's used to determine whether a piece of text contains some of the subjective information or not. And if there existany subjective information then it expresses whether the attitude behind the text is positive, negative or neutral. Understanding the opinions of other people will be of great help for commercial as well as for political use and also among others fields. Sentiment analysis, sometimes also known as "opinion mining" will let us know if there is been any change in public's opinion towards any of the aspect's regarding to business or any other areas. By obtaining the customer's feedback over the business regularly, one can improve the quality of service if required. Therefore, Sentiment analysis is said to be not a once done effort. In this paper, we have carried out a survey on Sentiment analysis, it's application's, challenges facing it, available tools and techniques implemented in this area.

Keywords: Sentiment analysis, Social Media, Sentiwordnet, StanfordCore, WEKA,Naive Bayes.

I. INTRODUCTION

Sentiment analysis also known as opinion mining is a process by implementing which we can analyze people's opinions, sentiments, evaluations, attitudes, and emotions.[1]It is one of the most active area which is used and implemented among the different research fields like natural language processing and it is also widely studied in data mining, Web mining, and text mining.Now a days Sentiment analysis has become a very important part in our day to day life therefore, the growing importance of sentiment analysis coincides with the growth of an individual along with social media's such as reviews, forum discussions, Twitter and also includes many other social networks.In order to take any decision sometimes we often seek out the opinions of others. This is true not only for individuals regarding to their personal or professional life but also it includes different business organizations and various other fields where we seek out for others opinion[1][2].Therefore, Sentiment analysis is said to be a process which is being implemented in almost every business and social field because, opinions are said to be the key factor to almost allhuman activities and are also said to be the key feature of human behaviors.Understanding the mood of humans is said to be very useful in many of the instances.Sentiment of any particular type of an entity can be very easily determined using different types of classification techniques, which are as described in below part of the paper. The advancement which have been made in some of the social networking medias like YouTube, has provided an opportunity for people in order to share their opinions in the form of audio, video, image, as well as in the form of text. Therefore,

we can say that sentiment analysis can be implemented not only on text input, but also it can be done on audio, video as well as on the image inputs.

Figure 1a. Sentiment Analysis

II. APPLICATIONS

Sentiment analysis, which plays a major role in today's technological world,has beenimplemented vastly in different fields. Therefore we can say that the applications for sentiment analysis are endless. [3]More and more we use the concept of sentiment analysis mainly in the field of social networking areas. However, apart from the social networking areas we use it in business analytics and also in certain situations where different forms of textsare needed to be analyzed, for the sentiment in it. In today's world Sentiment analysis is in demand because of its efficiency. By using the concept of sentiment analysis thousands of text documents can be processed at a time in order to detect for the sentiment in it.This particular process can be carried out in seconds of time rather than taking hours to complete it, by including the group of people who can manually carry out the complete process without any interruption[3][4]. Hence,due to its efficiency sentiment analysis has been utilized by many of the business organization in order to improve their quality of services.

The following are some of the areas where sentiment analysis is used and implemented:

1. BUSINESS

Sentiment analysis can be used in the field of business in order to provide the valuable insights representing that how people feel about the product brand as well as the service's provided by that particular business organization.In this case sentiment Analysis is said to be very helpful for the business organization in order to know and to identify the quality of their own products, which is determined based on the customer's feedback, where the feedback of the customer can either be positive, negative or neutral.

Therefore the main objective of using sentiment analysis in the field of business, by a particular organization is to find out the better ways in order to improve their quality of current services and to focus on negative feedbacks in order to work on it.

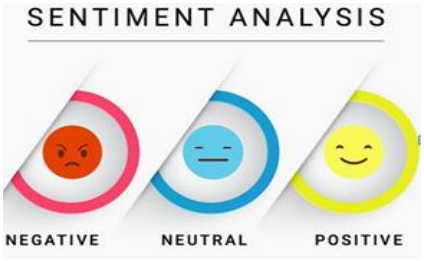

Figure 1b. Represents Different Opinions of People

2. SOCIAL NETWORKING MEDIAS

Apart from the business field sentiment analysis is mainly used in social networking areas in order to identify the spikes in sentiment (i.e., used to classify among the different opinions shared by people on the social media's). People now a day's expresses their feelings (i.e.,either positive or negative) regarding to government or any other organization or which can also be related to their personal life through the social networking sites.

Therefore, sentiment analysis plays a key role in classifying the different opinions of people that are expressed through social media's.

Figure 1c. Sentiment Analysis in the Field of Social Networking Areas

3. MARKETING

The very next important field where sentiment analysis implemented is in the field of marketing. In this particular fieldsentiment analysis is majorly used by the group of people who have involved themselves in the marketing area, where they implement the concept of sentiment analysis in order to determine the need of a customer by analyzing which they can manufacture new product's, that satisfies the customer needs. And also they gather different opinions and feedback from the people in order to improve the quality of product and their service.

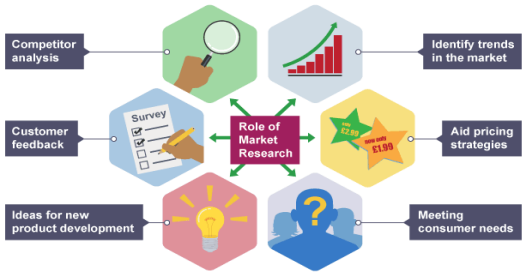

Figure 1d. Implementation of Sentiment Analysis in Marketing Field

4. ON-LINE ADVERTISING

Online advertising, where in case of social media's, an organisation may place an advertisement which could be favourable and provide a good review of a product but sometimes in order to promote their product they may de-promote the other products by giving a bad reviews over the other products.

Therefore, the major concept involved or used here is sentiment analysis where people are going to analyse which product to be bought based on the reviews given to the products.

5. MEASURING CUSTOMER SATISFACTION

When a particular product is developed or manufactured by an organization, the main intention behind this will be to satisfy the customer need. Any particular product or an item will be developed based or depending on the customer's needs and feedback.Hence, sentiment analysis is the major concept used by any organization in order to measure the level of customer satisfaction regarding to their products.For example, A Company may make an update to the product which is manufactured by them.By measuring the reactions of people through the process of sentiment analysis, the company can determine whether their customers took the update positively or negatively i.e., they determine whether the updated product will satisfy the customer needs or not.

III. CHALLENGES

Sentiment analysis is a particular term used for representing the large number of opinions and some of the tasks related to it, where each of the particular tasks go through their own unique challenges[6].

Firstly, the major challenge associated in the field of sentiment analysis is that, people who go through the sentiment analysis approach should be of open minded about their findings. As we know that sentiment analysis is a process, after going through which one can analyze the emotions as well as the thoughts of others which may sometimes hurt one's personal feelings or emotions. Therefore, one must be strong enough to take all opinions and feedback from people, which may be either positive or negative.Sentiment analysis, sometimes may be referred to as unambiguous process, but by having a closer view sometimes it becomes possible in order to identify or to determine how sentiment can be associated with any of the entities like the speaker or writer, the reader or listener, or any other entities.

Sometimes there may arise several situations where it becomes more difficult to determine whether the sentiment of the speaker is as same as in the situation or not. Consider the following statement given by a speaker;

Statement:"The dancer died due to a massive heart attack"!

The statement above represents a negative event i.e., death of a person, but by this statement it's not clear that whether the speaker is saddened by the event or not, i.e., we cannot identify the sentiment of speaker by referring to above statement.

Sometimes it becomes very difficult in order to analyze the sentiment of people in such particular situation when people react differently over the same utterance, where it becomes very hard to conclude the sentiment associated with respect to that particular situation.

Example: Consider a debate competition, on given a particular topic the teams will give different opinions over the same topic regarding to its pro's and con's which finally becomes hard to identify the sentiment associated with that topic.

Apart from the above challenges there exist another key challenge which is associated with sentimental analysis and it's termed as Named Entity Recognition – where the challenge associated here is to identify what the person is talking about.Example: Consider the following statement: "We watched the movie and went to dinner; it was awesome!"

In the above statement it's not specified that for what the term 'it' refers to(i.e., it's not clear that whether the term awesome is mentioned for the movie or for the food). Such type of challenge associated with respect to sentiment recognition, in case of sentiment analysis is referred to as "named entity recognition problem".

IV. TOOLS

The choice of which tool to be used and at what situation it should be used will be based on or it will be depended upon the specific problem that we are dealing in the field of Sentiment Analysis [7][8].There are many tools that are available for sentiment analysis among which some of them are mentioned as follows:

SENTIWORDNET: It's is one of the tool used in sentiment analysis for sentiment classification and it is completely based on the quantitative analysis approach and is freely available for research purposes, as well as it is included with a Web-based graphical user interface[GUI].

Itis a type of lexical resource which is widely used for opinion mining and it mainly represents three sentiment scores: i.e., positivity, negativity, and objectivity [7]. The main feature of this tool is that it has a Web-based GUI, and it is freely available for research purposes.

WORDSTAT: Word Stat is one of the particular tool which is used and implemented in the field of sentiment analysis [7]. This tool includes more than 9164 negative word patterns and 4847 positive word patterns. Even though it contains these number of word patterns, the sentiment is not measured using these word patterns.

Negative sentiment is measured by using the following two rules i.e.,1)The first rule is negative words cannot be preceded by a negation within three words in the same sentence. 2) The second rule is that the positive words can be preceded by a negation within three words in the same sentence. The rules for the positive sentiment are the same: positive words cannot be preceded by a negation as well as negative terms following a negation.

STANFORDCORENLP: When we need any part of the speech categories, as well as any syntactic

analysis i.e., the phrase structure OR dependency analysis and also any part of the named entities in the form of text i.e., in the textual format, this type of tool is implemented in order to carry out the process mentioned above [7]. These type of tools are been used as potential features by the sentiment analysis research community.

WEKA:WEKA is one of the very important tool that used in the field of sentiment analysis.If we already have some amount of data where each of the data content is associated with a key feature then we can use a tool named WEKA for the purpose of clustering these parts of the data [7].
This tool is said to be very highly configurable and it's easy to use the GUI availability.

NLTK: If one knows python, which is one of the most important programming language used now a days then NLTK is said to be a very smart choice as it includes the functionalists of the above two tools ie STANFORDCORENLP and WEKA [7].

Other than this, using NLTK one can very easily implement some of the lexical resources like WordNet which is often required and used in the field of sentiment analysis.

V. TECHNIQUES

There are many ways available in order to implement Sentiment Analysis. Ultimately, sentiment analysis is said to be a type of text classification problem which can be mainly divided into two areas[9][10]:

A) Supervised Learning

This is one of the technique used in sentiment analysis, which is mainly used for the construction of a "Classifier". The particular type of Classifier which is constructed is said to be responsible for categorizing the texts into either a positive, negative or neutral polarity.The three main classification techniques that come's under supervised learning are as follow

- Naïve Bayes
- Maximum Entropy
- Support Vector Machines (SVM)

1. NAÏVE BAYES CLASSIFIER (NB)

The Naïve Bayes classifier is one of the very simplest as well as the most commonly used classifier. The Naïve Bayes classification model computes the probability of a class, based on the distribution of the words in the document. It uses Bayes Theorem to predict the probability, that a given feature set belongs to a particular label.

$$P(label \mid features) = \frac{P(label)*P(features \mid label)}{P(features)}$$

The Naïve Bayes theorem was discovered by an English Presbyterian and mathematician named as **Thomas Bayes**. *P*(*label*) is the prior probability of a label. *P*(*features*/*label*) is the prior probability which represents that a given feature set is being classified as a label. *P*(*features*) is the prior probability which represents that a given feature set is occurred. If we consider the Naïve assumption which states that all features are independent, then the equation above can be rewritten as follows:

$$P(label \mid features) = \frac{P(label)*P(f1 \mid label)*\ldots\ldots*P(fn \mid label)}{P(features)}$$

2. Maximum Entropy Classifier (ME)

The Maximum Entropy Classifier (ME) is also known as a Conditional Exponential Classifier. It converts the labelled feature sets into vectors, using the encoding technique. This encoded vector is then used in order to calculate the weights for each feature which is then combined to determine the label for a feature set. This classifier is parameterized by a set of *X*{*weights*}, which is used to combine the joint features that are generated from a feature-set by an *X*{*encoding*}. Particularly, the encoding maps each *C*{(*featureset*, *label*)} is paired to a vector. The probability of each of the label is then calculated using the following equation:

$$P(fs \mid label) = \frac{dotprod(weights\ ,\ encode(fs\ ,\ label))}{sum(dotprod(weights\ ,\ encode(fs\ ,l))\ forlinlabels)}$$

3. Support Vector Machines Classifiers (SVM)

The major use of implementing the SVM classifier is to identify the linear separators, such that these of the separator's are used to separate the different classes.Example: consider the figure present below, there are 2 classes ie *x*&*o* and there are 3 hyper planes A, B and C. Hyper plane A provides the best separation between the classes, because the normal distance among any of the data points is the largest, so it represents the maximum margin of separation.

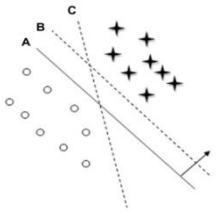

Figure 1e: Using Support Vector Machine on a Classification Problem

B) Unsupervised Learning: Unsupervised Learning consists of followingthree steps:

1. Implement thePOS tagging (Part of Speech), and then the two consecutive words are extracted in order to identify if their tagsconform to the given patterns or not.
2. Estimate the sentiment orientation of the extracted phrase.
3. Compute the average sentiment orientationof all phrases that were extracted in terms of positive or negative.

VI. CONCLUSION

Sentiment Analysis is the study of people's opinion, emotion or attitude towards an event, conversation on topics in general. Sentiment analysis which is also known as opinion mining is used in various applications, we use it to determine the mindset of humans based on their conversations with each other, in order to know the insights of one's business i.e., their quality of service and their products which is determined based on the customers feedback, When applied to social media channels, it can be used in order to identify the spikes in sentiment of the people, who expresses their opinion and feelings through social media channels.Sentiment analysis is a particular term used for representing the large number of opinions as well as some of the tasks related to it, where each of the tasks has its own unique challenges. In order to carry out the process of sentiment analysis in a perfect, as well as in a structured manner we use a number of tools in this field among which some of the tools are as mentioned and explained in the above part of the paper. Sentiment analysis, which is rather referred to as a text classification problem, utilizes many of the techniques in order to implement the process and to overcome the text classification problem.Sentiment analysis is done not only on text inputs, but it can also be implemented on audio, video and image inputs.

VII. ACKNOWLEDGEMENT

We are using this opportunity to express our gratitude to thank all the people who have contributed in some way to the work described in this paper. Our deepest thanks to our project guide Mr Hemanth Kumar A, Asst. Professor Dept. of Computer Science and Engineering, GMIT, Bharathinagar for giving timely inputs and giving us intellectual freedom of work.

VIII. REFERENCES

1. Pang, B., & Lee, L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts.
2. WMedhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering journal, vol. 5, pp. 1093–1113, 2014.
3. Sentiment Analysis: Overview, Applications and Benefits-By Mikaela Berman, July 5, 2017 Categories: Big Data & AnalyticsTags: big data, data analytics, Sentiment Analysis

4. Sentiment analysis algorithms and applications – Science direct- https://www.sciencedirect.com/science/article/pii/S2090447914000550
5. https://www.quora.com/what-are-the-applications-of-sentiment-analysis
6. www.adweek.com/digital/5_key_challenges_in_sentiment_analysis/
7. Basic tools used for Sentiment Analysis algorithms https://scholar.google.co.in/scholar?start=90&q=basic+tools+for+sentiment+analysis+algorithms&hl=en&as_sdt=0,5
8. Sentiment Analysis Tools to Track Social Marketing Success |- iProspect
9. Sentiment analysis algorithms and applications: A survey- https:// www.sciencedirect.com/science/article/pii/S2090447914000550
10. https://www.nowpublishers.com/article/Details/INR-011