ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 5, May 2017

Text Based Emotion Detection Techniques

[1]Akshaya Chandorkar, [2]Asha Bharambe
[1] P. G. Student, Vivekanand Education Society's Institute of Technology
[2] Assistant Professor, Vivekanand Education Society's Institute of Technology
[1]akshayachandorkar29@gmail.com, [2]asha.bharambe@ves.ac.in

Abstract: **— Emotion can be expressed by different forms such as written text, speech or video. With the growth of social networking, textual data has proven to be the main tool of the interaction. Emotion Detection in text documents is essentially a content-based classification problem which has evolved from the domains like Natural Language Processing as well as Machine Learning.**

Computational analysis of emotions has been considered a challenging and interesting task. However, there are few prior works who work with textual input to analyse these emotions. This paper explains different techniques of emotion detection, challenges faced and various applications of emotion detection. Also, it presents the implementation of the system.

Index Terms: **— Emotion Detection, Sentiment Analysis, Machine Learning, Naïve Bayes, SVM.**

I. INTRODUCTION

Emotions are the colors of the soul. Emotion, in everyday speech, may refer to the affective aspect of consciousness, a state or feeling, or a conscious mental reaction towards an object accompanied by behavioral and or physical changes. Micro-blogging websites have evolved to become a source of opinionated information. Twitter is popular micro-blogging site which generates 500 million tweets per day as of November 2016. Because almost all the tweets are public, these rich data offer new opportunities for doing research on data and text mining.
She is lost after her father's death.

Going beyond a simple negative sentiment, the above statement is an indicator of sadness. Classification into deeper emotion helps us to understand how such conversations develop, how people influence one another through emotional expressions, and how news is shared to elicit certain emotional reactions.

II. MOTIVATION

With the rise in social networking, people tend to express and share opinions with each other on the social media platforms like twitter, facebook, gmail etc. The communication is constantly evolving towards the goal of making it more human and real. The main motivation behind this work is to investigating possibilities of detecting emotions by multiple classes in short texts like twitter and to analyze the difficulties and drawbacks in the same rather than just classifying them as positive or negative sentiment.

III. SENTIMENT ANALYSIS

A. Sentiment analysis

The sentiment is the ripened fruit of fantasy. The process of identifying and categorizing opinions expressed in a piece of text, to determine whether the writer's opinion towards a particular topic, product, etc. Is positive, negative, or neutral.

It is the process of determining the emotional tone behind a series of words. Sentiment analysis is very much useful in social media monitoring as it permits us to gain an overview of the wider public opinion behind certain topics. [1]

B. Sentiment analysis vs emotional analysis

Sentiment analysis is the process of classifying a piece of text into polar classes. From machine learning perspective, sentiment analysis is a two or three class classification problem known as polarity classification or sentiment classification.

On the other hand, emotion recognition is a more recent and emerging field in opinion mining which tries to expand boundaries of sentiment analysis wider in hope to

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 5, May 2017

gain better understanding of people's textual opinion; on these approaches, they have a micro-level, multi-class (anger, disgust, fear, happiness, sadness, love, excitement, shame and surprise) glance at public's opinion.

Emotion recognition is a special case of sentiment analysis. The output of sentiment analysis is manufactured in terms of either polarity (e.g., positive or negative) or in the form of rating (e.g., from 1 to 5). Emotions are a more detailed level of analysis in which the result are depicted in more expressive and fine-grained level of analysis. Sentiment analysis deals with only text, while emotions can be expressed by text, images, audio, video, facial signs etc.

C. Moving towards emotion analysis

Sentiment analysis (sa) analyses online texts and states the writer's sentiments or opinions towards a particular object or any of its subsumed features or aspects. In order to evaluate sentiments in sentiment analysis, we need to detect terminologies, called opinion carrying words or simply opinion-words, which are the base for sentiment analysis to perform the evaluation.

This evaluation is insufficient for an accurate and a more detailed evaluation, though. That is because sentiment polarity (positive and negative) does not fetch the affective meaning that writers give to an item or to any of its related features. Therefore, there is a need for a stronger and more fruitful evaluation that is able to show writers' opinions at the emotional level of evaluation towards what he/she writes.

Emotions detection from text provides a strong and expressive opinion evaluation over conventional approaches. Emotion usefulness comes due to its importance in the academic community, in industry, and in a variety of applications such as politics, marketing, education, and much more. [2]

IV. EMOTION DETECTION

A. Emotion

Emotion is energy-in-motion. It is a way by which one can express himself/herself in life. Emotions are a crucial part of human interaction. Textual data is of great importance to researchers. It requires small storage medium so, textual data is the most proper medium for network transmission. Now-a-days, most of the communication is performed through text. Basically, the emotions are divided into two broad types i.e. Positive and negative. There are six basic categories of emotions stated by ekman namely happiness, sadness, anger, disgust, surprise and fear. Examples of positive emotions include happiness, laughter, interest, joy, love and so on. Negative emotions include anger, fear, sadness etc.

According to psychologist robert plutchik, there are eight basic emotions: joy, trust, fear, surprise, sadness, anticipation, anger, and disgust. Plutchik created the wheel of emotions, which illustrates the various relationships among the emotions. As we stated earlier, plutchik's eight basic emotions are joy, trust, fear, surprise, sadness, anticipation, anger, and disgust. Each primary emotion also has an opposite emotion, so that:

- Joy is the opposite of sadness.
- Fear is the opposite of anger.
- Anticipation is the opposite of surprise.
- Disgust is the opposite of trust.

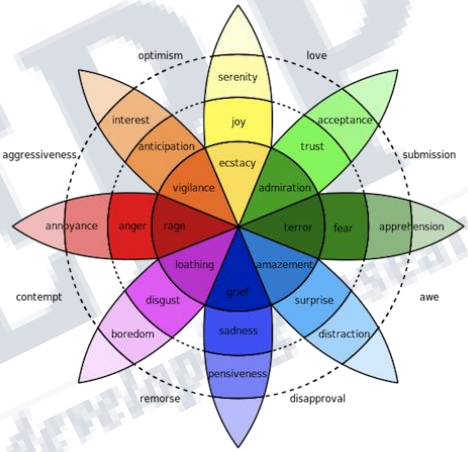

Fig. 1: plutchik's wheel of emotion, adapted from [3]

The intensity of the emotion is indicated by the color. The darker the shade, the more intense the emotion. For example, trust at its least level of intensity is acceptance. At its highest level of intensity, trust becomes admiration.

B. Applications of emotion detection Intelligent tutors

Now-a-days, several automatic tutoring systems have been deployed for different educational applications. For these tutoring systems to be affective teachers, it is important that they should possess skills of a traditional teacher. One of these skills is to understand how well the class is following you. So, the solution is to implement a tutoring system which deals with the emotional state of the students too. As it is tutoring system, emotion labels can be neutral, confusion, boredom, frustration, delight, flow and surprise.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 5, May 2017

Internet chat

Since chat applications are very popular and commonly used, to know the emotion of opposite person, it is essential and useful to perform emotion analysis of chat messages.

Mental health monitoring [4]

Mental health issues pose risks to lives and wellness of million people. In times of social media, users often use media like twitter and facebook to express their emotion. By using ea, one can detect the emotion from posts and predict the illness if there is any. E. G. From the employee's tweet "suffering from fever", boss can predict his absence of next day using ea.

Marketing

Emotions are the main drivers of most human behavior for purchasing decisions. Emotion analysis helps marketers to advertise according to the current emotion of the person. It will also help recommender system to recommend the products based on the emotions.

Video games [5]

Video games belong to the wide area of entertainment applications. So, emotion analysis will help the developers to make player tightly attached with the game. Sometimes long, repeated, frequent stimulation may cause a negative reaction of the player so, emotion analysis helps developers to remove such stimulation.

V. TEXT BASED EMOTION DETECTION TECHNIQUES

A. *Keyword spotting technique*

The keyword spotting technique can be described as the problem of finding the appearance of keywords from a given text. This is the most naïve approach and probably also the most popular because of its accessibility and economy is spotting keywords from an emotive vocabulary. [6] text is classified into affect categories based on the emotion words like "distressed", "sad", "happy", "love." The process of keyword spotting method is shown in the figure:

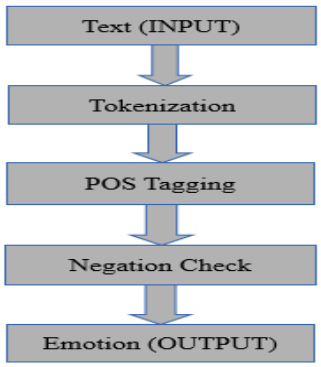

Fig. 2: keyword spotting technique

The method consists of five steps shown in the above figure where a sentence, text document or any tweet is taken as input and the output is generated as an emotion class i.e. Happiness, anger, sadness, fear, disgust.

Text (input): this is the very first step in which text data is gathered. Data can be a sentence or text document. Here we will take the example of a sentence."she is very beautiful."

Tokenization: tokenization is a way to split the text into tokens. These tokens could be paragraphs, sentences, or individual words. Tokenization of the above sentence would be-SheIsVeryBeautiful.

Pos tagging: in corpus linguistics, part-of-speech tagging (pos tagging or post), also called grammatical tagging or word-category disambiguation, is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on its definition and its context. A simplified form of this is commonly taught to school-age children, in the identification of words as nouns, verbs, adjectives, adverbs, etc. After pos tagging, the above sentence will be-

She|prp is|vbz very|rb beautiful|jj .|.

Negation check: once pos tagging is done, negation can be checked. The sentence is checked whether any kind of negative word like not, neither, never etc. Is present in it or not. The above sentence is not containing any kind of negative word. If the sentence would have been "she is not very beautiful," this step would have resulted-

Not

As the negation word and emotion would have been changed. I.e. Opposite.

Emotion: finally, an emotion class will be found as the required output. Here, happy is the word indicating emotion so that the final result will be-

Happiness

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 5, May 2017

B. *Lexical affinity method*

Keyword spotting method is an easy to use and straightforward method. Lexical affinity method is slightly more sophisticated than keyword spotting. This approach is an extension of keyword spotting technique; it assigns a probability for a particular emotion to an arbitrary word apart from picking up emotional keywords. Consider an example,"i met my old classmate by accident". In the above sentence, the word "accident" is indicating high probability which having a negative emotion. But the exact situation in this sentence that accident word not showing negative emotional assessment. [7]

C. *Learning-based methods*

Initially, problem was to determine the sentiment from the given input text data but now the problem is to classify the input texts into different emotion classes. Learning-based methods are being used to analyze the problem in a different manner. In machine learning, we need large annotated data to train the machine. Unlike keyword-based detection methods, learning-based methods try to detect emotions based on the previously trained classifier, which apply theories of machine learning such as naïve bayes classifiers and support vector machine. [8]

Naive bayes classifier

The naïve bayes classifier is a simple, trained and probabilistic classifier model. It assumes that every feature or attribute of the dataset is independent of each other. So, it is named "naive". Multinomial naïve bayes model is a specialized version of naïve bayes model that is specially designed for text documents.

Pseudo code for naïve bayes classifier:

Step 1: collect all the unique words from document.

Step 2: represent each document by a vector of words. I.e. Convert all documents into feature set, where the attributes are possible words, and the values are a number of times a word occurs in a given sentence.

Step 3: calculate the probability of each class. The probability of class or prior probability is given by a number of documents or text having the class c to the total number of documents.

$P(c) = n_c / n$

Step 4: calculate the probability of the word being in class. The probability of the word being in class or conditional probability is given by a count of the word in the class plus 1 smoothing to total words in the class plus vocabulary count.

$P(w \mid c) = count(w,c) + 1 / count(c) + |v|$

Step 5: classify the new sentence. Probability of new doc being in class c is proportional to the prior probability of class c multiplied by the conditional probability of every word in the new document.

P(c | new) = p(c) * conditional probability of every word in the new document.

Support vector machine classifier

Support vector machine is another popular classification algorithm. It is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. Svm is primarily a classier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels.

It is explicitly used to find the best separating line. It searches for the closest points which are called "support vectors". Once it has found the closest points, it draws a line connecting them then svm will declare the best separating line that bisects. Svm supports both regression and classification tasks and can handle multiple continuous and categorical variables.

D. *Hybrid method*

The hybrid method is the combination of any two methods. Since keyword-based methods can't predict emotion if there is no any emotional word and learning-based methods alone gives less accuracy, some systems use hybrid approach by combining two techniques which will help to improve the accuracy. [9]

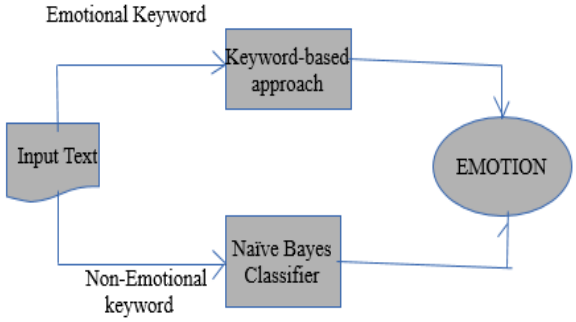

Fig. 3: the hybrid approach

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 5, May 2017

If input sentence has emotional keywords, we apply the keyword-based approach to the system since the keyword based approach still shows high accuracy for the sentences having keywords. In other cases, the system will recognize emotions from the sentence with no emotional keywords using machine learning method, naïve bayes classifier.

E. Challenges

Following are the major issues where intensive research is required.

Ambiguity in keyword definitions

Using emotion keywords is a straightforward way to detect associated emotions, the meanings of keywords could be diverse and indefinite, as most words could change their meanings according to different usages and contexts. E.g. "i avoided an accident" and "i met her by accident" are the two sentences having same word accident but in the first sentence it is having negative polarity but in the second sentence, the same word is indicating some situation which has happened unexpectedly.

Incapability of recognizing sentences without keywords

The keyword-based approach is totally based on the set of emotion keywords. Therefore, sentences without any keyword would imply that they do not contain any emotion at all, which is so wrong.
E.g. "come here and sit". Is the sentence which does not show any emotion so such sentences can't be classified.

Lack of linguistic information

Syntax structures and semantics also have influences on expressed emotions.
E.g. "i ignored him" and "he ignored me" would suggest different emotions from the first person's perspective. As a result, ignoring linguistic information also poses problems to emotion detection.

Difficulties in determining emotion indicators

Now-a-days, most intuitive features may be emoticons which can be seen at social media very frequently. So, there is a challenge to determine the emotion in the text format from the emoji.
e.g. I am 😂
The above emoji is for extreme happiness which causes crying. The challenge is to identify the meaning of the emoji.

Multiple opinions in a sentence

A single sentence can contain multiple opinions along with instinctive and truthful portions. It is helpful to isolate such clauses. It is also important to estimate the strength of opinions in these clauses so that we can find the overall emotion in the sentence.
E.g. She is very talented at her acting but she is totally ugly in looks.

Comparative sentences

A comparative sentence expresses a relation based on similarities or differences of more than one object. Research on classifying a comparative sentence as opinionated or not is limited. Also, the order of words in comparative sentences manifests differences in the determination of the opinion orientation.
E.g. The sentence, camera a is better than camera b communicates a completely opposite opinion from camera b is better than camera a.

Sarcasm

Sarcasm and irony are quite difficult to identify. Sarcasm is very often used in social media.
E.g. Earth is full. Go home. Basically, it will give us the emotion like neutral or some amount of sadness is there but actually, it means to go to hell. [10]

Use of abbreviations and acronyms

As tweets consist of short messages expressed within the limit of 140 characters, the language used to express emotions in tweets differs significantly from that found in longer documents. Language use on twitter is also typically informal [11]. It is common for abbreviations, acronyms, emoticons and misspellings to occur in these short messages. To detect emotions from such language becomes a challenging task.
E.g. "i m grt". This sentence should have written as "i am great". In this example, 'am' is written as 'm' and 'great' is written as 'grt'.

VI. IMPLEMENTATION

Implementation Has Been Done Using Keyword Spotting Method And Learning Based Method To Analyze The Accuracy.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 5, May 2017

A. Datasets

Two datasets have been used in the experiments reported in this paper.

Dataset 1

Dataset 1 contains 7516 instances with number, emotion, id & text attributes. The number is the serial number. Emotion contains labels of emotion such as happiness, anger, fear, sadness, disgust. Id is the number assigned to every emotion. And text contains tweets which we have to classify.

	A	B	C	D	E	F	G	H	I
1	NUMBER	EMOTION	ID	TEXT					
2	1	happiness	1	On days when I feel close to my partner and other friends. When					
3	2	fear	3	Every time I imagine that someone I love or I could contact a seric					
4	3	anger	2	When I had been obviously unjustly treated and had no possibilit					
5	4	sadness	4	When I think about the short time that we live and relate it to the					
6	5	disgust	5	At a gathering I found myself involuntarily sitting next to two pec					
7	6	disgust	5	When I realized that I was directing the feelings of discontent wit					
8	7	sadness	4	I feel sadnessy when when I realize that I consider material thing:					
9	8	happiness	1	After my girlfriend had taken her exam we went to her parent's p					
10	9	fear	3	When, for the first time I realized the meaning of death.					

Fig. 4: dataset 1

Dataset 2

Dataset 2 contains 206 instances with emotion, date, name, username & tweet attributes. Emotion contains labels of emotion such as happiness, anger, fear, sadness, disgust. The name is the name of the user. Username is the username of the user on twitter. And tweet contains tweets which we have to classify.

	A	B	C	D	E	F	G	H
1	Emotion	Date	Name	Username	Tweet			
2	happiness	Mon May	lebron	mtgillikin	i love lebron. http://bit.ly/PdHur			
3	happiness	Mon May	lebron	princezzc	@sketchbug Lebron is a hometown hero t			
4	happiness	Mon May	lebron	peterlike	lebron and zydrunas are such an awesome			
5	happiness	Mon May	lebron	emceet	@wordwhizkid Lebron is a beast... nobod			
6	happiness	Mon May	iphone ap	CocoSava	downloading apps for my iphone! So muc			
7	happiness	Mon May	visa	Dreambig	good news, just had a call from the Visa of			
8	disgust	Thu May 1	"booz alle	JoeSchuel	Booz Allen Hamilton has a bad ass homegr			
9	fear	Sat May 1(	google	vamsmack	Played with an android google phone. The			
10	disgust	Sat May 1(	itchy	MarissaLe	omg so bored & my tattoooos are so			
11	disgust	Sat May 1(	itchy	robloposk	I'm itchy and miserable!			

Fig. 5: dataset 2

B. Implementation

Keyword spotting method

The framework consists of two modules, namely:

A) Preprocessing for data cleaning: sentence contains many undesirable symbols, links etc. That may compromise with efficiency and effectiveness of the algorithm applied. Thus, processing of data is required. It contains removal of hashtags, symbols, hyperlinks etc.

B) Scoring module: after preprocessing the sentence, we are calculating the semantic score of the opinion carriers. I.e. The adjectives, verbs, and adverbs. I have used the corpus for adjectives, adverbs and verbs values from the reference paper [12]. The table shows few tuples of adjectives values:

Table I: Adjective Emotion Value Vector

WORD	Happiness	Anger	Sadness	Fear	Disgust
abduction	1.09	3.54	3.93	4.11	3.47
accident	1.06	2.82	3.61	3.69	1.88
alone	1.27	1.84	3.65	3.30	1.56

Adverbs are not so very important but they can change the meaning of the adjective which may convey positive or negative effect. Therefore, adverb strength is taken into consideration. Therefore, adverb strength is taken into consideration. We consider same values as assigned to the most frequently used verbs and adverbs defined by kumar and sebastian described in the table.

Table ii: verb and adverb strength

Verb	Strength	Adverb	Strength
Love	1	complete	+1
Adore	0.9	most	0.9
Like	0.8	totally	0.8
Enjoy	0.7	extremely	0.7
Smile	0.6	too	0.6
Impress	0.5	very	0.4
Attract	0.4	pretty	0.3
Excite	0.3	more	0.2
Relax	0.2	much	0.1
Reject	-0.2	any	-0.2
Disgust	-0.3	quite	-0.3
Suffer	-0.4	little	-0.4
Dislike	-0.7	less	-0.6
Detest	-0.8	Not	-0.8
Suck	-0.9	never	-0.9
Hate	-1	hardly	-1

Emotion scoring

Algorithm for emotion calculation

Step 1: if an adverb or verb is encountered after another adverb or verb then we add it. We repeat this process till another adjective isn't encountered.

Step 2: now if the added value of adverbs or verbs is less than 0 i.e. Negative, then for the upcoming adjective, we subtract its value from 5 instead of multiplying.

Step 3: and if added value of verbs or adverbs is positive and >=0.5 then multiply it with the upcoming adjective else multiply 0.5 with an upcoming adjective.

I have extracted few tuples of isear dataset and tested the accuracy of my system. The system is giving 100% accuracy.

Learning-based methods

I have implemented support vector machine and naïve bayes classifier using orange tool. Orange (http://orange.biolab.si) is a general-purpose machine learning and data mining tool. It features a multilayer architecture suitable for different kinds of users, from inexperienced data mining beginners to programmers who

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol4, Issue 5, May 2017

prefer to access the tool through its scripting interface. It is open source and component based tool packed with features of data analytics. Orange widgets provide a gui to orange's data mining and machine learning methods. They include widgets for data entry, classification, regression, evaluation, visualization etc.

I have used two datasets to implement classification.

C. Result

Keyword spotting method gives 100% accuracy if emotional keywords are there. Svm gives 97.9% and naïve bayes gives 84.9 % accuracy for dataset 1 whereas for dataset 2, svm gives 48.5% and naïve bayes gives 63.1% accuracy.

Result of learning-based method is-

Dataset / Technique	DS2	DS3
Support Vector Machine(SVM)	97.9%	48.5%
Naïve Bayes Classifier	84.9%	63.1%

Fig. 6: result analysis of learning based methods

V. CONCLUSION & FUTURE WORK

Communication in the form of text is the most accustomed approach. The emotion detection from text aims to find the particular emotion by studying the input text. In this paper, different techniques for finding out emotion are studied. After the survey, it is concluded that keyword spotting method gives 90% accuracy in case of sentences with emotional keywords but there is a problem when the emotional word is not present. In such cases, hybrid method (combination of any two methods) gives more accuracy as compared to all other methods.

Text based emotion detection has bright future. Despite effectiveness in various applications, still, there are challenges that should be taken into consideration for further improvement.

REFERENCES

[1] bing liu, "sentiment analysis and opinion mining", morgan & claypool publishers, vol. 168, pp. 7-12, may 2012.
[2] kristian bannister, "understanding sentiment analysis: what it is & why it's used", https://www.brandwatch.com, january 2015.
[3] robert plutchik, "the nature of emotions", vol. 89, pp. 344-350, july-august 2001.
[4] vaibhav tripathi, aditya joshi, pushpak bhattacharyya, "emotion analysis from text: a survey", 2016.
[5] a. Landowska, m. Szwoch, w. Szwoch, m.r. Wróbel, a. Kołakowska, "emotion recognition and its applications", human-computer systems interaction: backgrounds and applications 3, advances in intelligent systems and computing volume 300, 2014.
[6] shiv naresh shivhare and prof. Saritha khethawat, "emotion detection from text", conference: data mining and knowledge management process (dkmp 2012), at new delhi, may 2012.
[7] chetan r. Chopade, "emotion text based recognition: a survey", international journal of science and research (ijsr) issn (online): 2319-7064, vol 4, june 2015.
[8] uma nagarsekar, priyanka kulkarni, aditi mhapsekar, dr. Dhananjay r. Kalbande, "emotion detection from 'the sms of the internet'", ieee recent advances in intelligent computational systems (raics), 2013.
[9] yong-soo seol and dong-joo kim, "emotion recognition from text using knowledge-based ann", the 23rd international technical conference on circuits/systems, computers and communications (itc-cscc 2008), 2008.
[10] mr. Nilesh m. Shelke, "approaches of emotion detection from text", international journal of computer science and information technology research issn 2348-120x (online), vol.2, issue 2, pp:(123-128), month: april-june 2014.
[11] jacob eisenstein, "what to do about bad language on the internet", proceedings of naacl-hlt 2013, association for computational linguistics, pages 359–369, 9–14 june 2013.
[12] akshi kumar, prakhar dogra, vikrant dabas, "emotion analysis of twitter using opinion mining", institute of electrical and electronics engineers, august 20, 2015.