Issues in Sentiment analysis

[1]Prof. Miss. D. S. Ulhe, [2]Prof.Miss Sonali V. Muddalwar

[1, 2] P.G. Department of Computer Science & Technology (DCPE) Shree Hanuman Vyayam Prasarak Mandal, Amravati (Maharashtra)

Abstract: *Natural language processing is concerned with artificial intelligence and linguistics. It provides opportunity to interact with computers using human natural languages. There are varieties of applications of NLP. The AI and business intelligence contributes in extraction of meaning and knowledge represented in natural language text. The major areas are online media application, academic application, biomedical application etc. No application claims hundred percent NLP implementation. There are very many difficulties and issues in natural language analysis. In business application and AI, text analysis and text mining is used to extract information and knowledge from the text. Attribute extraction, author identification, text subject identification are some applications in these areas. The sentiment analysis to extract the emotions, feelings in the thought is difficult. This paper discusses the issues in identification of emotions and feelings in the text.*
Keywords: *Artificial Intelligence, business intelligence, Emotion extraction, text analysis, text mining*

I. Introduction

Natural language processing is concerned with artificial intelligence and linguistics. It provides opportunity to interact with computers using human natural languages. Over the last few decades, there is increasing body of research on extracting the human emotions from the text. In this study automatic classification of emotions like anger, disgust, fear, joy ness and sadness etc. in text have been studied on ISEAR(International Survey on Emotion Antecedents and Reactions) dataset. There are different techniques have been invented to extract emotion from the text, such as text mining, Empirical study, Emotion extraction engine, Vector Space Model ,Emotion Markup Language, etc. While analyzing the calculated results using any existing technique, we found that every new system overcome the problem of existing system but still not a single application claims hundred percent realistic results. And hence there is a need to discuss the issues regarding failure of existing system.

II. Need

As we have discussed earlier, via sentiment analysis it will be possible to get solution on real time system. Emotion extraction from facial expression, visual clipping is somewhat easy, but this is not sufficient. Textual material also has sentiment and most of the time there is need to extract emotions from the text. The previous research in this area claims that it can be used in next generation intelligent robotics, artificial intelligence and psychology etc. With this it can also be useful in various areas like criminal identification and authentication, subject identification etc. Hence it needed, to study all these issues together for getting solution on this matter. With this comparative study we can find out issues and some solutions.

III. Comparitive study of emotion extraction techniquies

3.1 Text Mining:-

Text mining, sometimes alternately referred to as *text data mining* roughly equivalent to *text analytics*, refers to the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interestingness. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, *sentiment analysis*, document summarization, and entity relation modeling (*i.e.* learning relations between named entities).

3.2 Empirical study

Empirical study covers the experimental study with formal problem definition, conceptual implementation, data, features and a note on parameter tuning. For formal problem definition machine learning model is used. In this model determination of emotion of linguistic unit can be cast as multicast classification problem.

In order to make the annotation process more focused, emotion is annotated from the point of view of text. While targeting primary emotions the sentences are also marked for other affective contents i.e. background mood, secondary emotions via intensity and textual cues. Disagreements in annotations are resolved by a second pass of tie-breaking by the first author, who chooses one of the competing labels. Eventually the completed annotations will be made available.

Table 1: Basic emotions used in annotations

Abbreviation	Emotion class
A	Angry
D	Disgusted
F	Fearful
H	Happy
Sa	Sad
Su+	Positively Surprised
Su-	Negatively Surprised

Using above table1 emotion have been extracted and then further percentages of annotated labels have been determine using following table2 (Ex. of Some text passages).

Table 2: Percent of annotated labels

A	D	F	H
12.34%	0.89%	7.03%	6.77%
N	Sa	Su+	Su-
59.94%	7.34%	2.59%	3.10%

After that this model has also worked on classifying the emotional versus neutral examples as shown in table 3.

Table 3: Emotional v/s neutral example

E	N
40.06%	59.94%

Further percentage of positive, negative and neutral emotions can be determined using following table4.

Table 4: Positive v/s negative v/s neutral example

PE	NE	N
9.87%	30.19%	59.94%

This model based on analytical calculations for determine the text to speech synthesis as goal.[1]

3.3   Emotion extraction engine
3.3.1 Background
        In human language, sounds and letters make up the significant part of the communication. Words are group of letters and may be combined in many ways to make up a sentence. [2] Sentences can be broken down into two groups of words i.e. function words and contents words. Function words include noun, adjectives, verbs and adverbs, while content words are prepositions, conjunctions and auxiliary words. Emotion extraction engine have decompose the sentences and analysed them in terms of Sentence structure and analysis, Declarative statements, interrogative statements, finite verb phrases and negative statement.
        An emotion extraction engine that can analyze user input in the form of text sentences has been developed. When emotional content is detected the engine will send the parameters needed for selecting the expressive images across the network. When receiving side engine receives the parameters, the corresponding expressive images will be selected and displayed. In this way, the transmissions of images are avoided since the only data transmitted over the network is the text parameters. As a result the bandwidth requirement is extremely low. [2]

The working flow of the emotion analysis system is shown in figure 1.
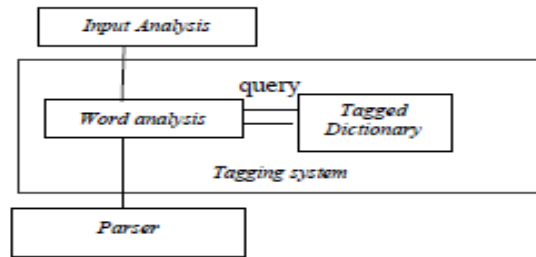
3.4 Vector Space Model

Vector Space Model(VSM) is widely used in information retrieval. Where each document is represented as vector and each dimension correspond to a separate term. VSM approach is used for human emotion recognition.(HER) from text. In this approach the effect of stemming and emotional intensity on emotion classification in text is measured. VSM based classification on short sentences can be as good as other well-known classifiers. [3]

3.5 Emotion Markup Language

The present draft specification of emotion markup language aims to strike a balance between practical applicability and scientific well-found ness. The language is conceived as a "plug-in" language suitable for use in three different areas: (1) Manual annotation of data;
(2)Automatic recognition of emotion-related states from user behavior; and
(3) Generation of emotion-related system behavior.
  Some tags used in EML
1.
2. <category>
3. <happy>:-)</happy>
4. <sad>:-(</sad>

IV.   Different issues
From comparative study of emotion extraction techniques different issues have been generated as listed below.
• As per the text mining technique we can only identify the subject of the text but we can not determine its approach i.e. positive, negative or neutral.
• With identification of text approach it is also needed to find out the intensity level of emotion i.e. low , moderate or extreme.
• It should be also possible to identify the contradiction between feelings.
• It should be also possible to detect the proper meaning of the word as lots of words have multiple meanings.
• It should be possible to extract proper results for multi-lingual text also.
• If there is a contradiction between actual emotions and the written text, then it cannot be possible to analyse the actual sentiments from the text.
• If the author does not have good collection of vocabulary then he will unable to express intensity and the actual thoughts in the text .
• If it has been implemented on internet then it must be possible that our search must show result according to sentiment and nature of the text.
• While implementing it in marketing application it must gives us comparative result of products.
• Lots of applications are designed and implemented for the internet users. Most of the times people using different symbols for representing different sentiments of user. It should be possible to extract emotion from symbolic text also.
• Some texts are machine generated such text contains meaning with emotions but machine has no sentiments so it generate contradictory results.

V.   Techniques to overcome pitfalls of sentiment analysis
The lots of work have been done on extracting emotions via Natural language Processing. Very many areas have been used for such implementations. Extracting emotions from heard audio or video is somewhat easier as compared to extracting sentiments from plain text.

Peoples have worked a lot in this area also but there is no system claiming hundred percent realistic results. Hence there are some hesitations or doubts while implementing it.

From the comparative study we can clearly determine the reasons behind the failure of acquiring hundred percent realistic results.

Every new system is better than existing one and provides solutions to overcome the limitations. When we go through that system it also has some limitations. We think that, this happens because of individual study of parameters. If all these issues can be investigated together then definitely we can proposed a better system which will be much more nearer to realistic result.

We knew that it is very difficult to study and combine implementation of all systems, but if it will be implemented then really it will gives lots of surprising results in IT world like next generation AI, accessing data from database via NLP, software development using NLP etc.

For acquiring the system which will discuss the issues stated as above. While developing such system if think on some parameters as listed below then it will be a good one.

- For determining the approach of text we have to extend the mining criteria.
- Once we will determine the definite approach of our text we should also work on determining its intensity. For eg. Once we determine the text is related to happiness so we can also extract the data from the text related to such emotions. After extracting this parameter it should be differentiated into the categories and calculating its comparative percentage we can deals with the intensity of the emotion.
- To overcome the contradictory results at preliminary level we must have to identify the creator of the text.
- For determining exact meaning of words that have multiple meanings we should maintain the databases for such words like dictionary and then also find out sentiments of other words related to that.
- For determining exact meaning of multilingual words. We should maintain the databases for such words.
- To overcome the contradiction between actual emotions and the written text, while analyzing the sentiment from the text if we will take snapshot of the author. Then from his facial expression we can determine the emotions from the text as per his facial expression.
- While searching on internet for showing result according to sentiment and nature of the text. It should be first find out the meaning & then fires the searching query.
- While implementing in marketing area before generating the results it should analyze the comparative study of follow up and then generate appropriate result.
- While extracting emotion from the text frequently used symbols can also be taken into consideration.
- The combine study of multiple modalities in video, audio, visual and text to improve the classification performance.[4]

The sentiment analysis to extract the emotions, feelings in the thought is difficult. This paper discusses the issues in identification of emotions and feelings in the text.

VI.   Conclusion

The sentiment analysis to extract the emotions, feelings in the thought is difficult by focusing on single parameter. We can conclude that if we want the better solution then we must have to work on different parameters and their approach as discussed in our paper.

References
[1]   Cecilia Ovesdotter Alm, Dan Roth and Richrd Sporoat on "Emotion from text: Machine learning for text based emotion prediction" from department of linguisics UIUC
[2]   Graddol , Cheshire and Swann: 'Describing language'
[3]   Taner Danisman and Adil Alpkocak: Feeler:'Emotion classification of text using vector space model
[4]   G. Mishne, Experiments with mood classification in blog posts. In: Style 2005-Ist workshop on stylistic Analysis of text for information access, at SIGIR(2005)