Sentimental Analysis and Prediction from Social Media using Text

Ravina Patil[1] , Rupali Pawar[2] , Varsha Jadhav[3], Daniyal Shaikh[4]
Department of Computer Engineering
[1] U.G. Student, SSBT's COET Bambhori , Jalgaon , Maharashtra , India

Abstract- *Detecting emotion from text is a relatively new classification task and advancements in textual analysis have allowed the area of emotion detection to become a recent interest in the field of natural language processing. There is still a question on how to detect emotion from a text input. To solve this problem, this project generates an Emotion Detection Model to extract emotion from text at the sentence level. The proposed methodology does not depend on any existing affect lexicons such as Word Net Affect. Our method detects emotion from a text-input by searching direct emotional key words from that input. To make the detection more accurate, emotion-affect-bearing words and phrases were also analyzed. The experiments show that the method could generate a good result for emotion detection from text input. To detect emotion from text we have considered Ekman's six emotions class (joy, sadness, anger, disgust, fear, surprise). Our approach showed above 77% accuracy in detecting emotion from text input*

Keywords- Emotion Detection, Lexicons

I. INTRODUCTION

Emotion is one type of affect, other type of being mood, temperament and sensation. Emotions have been widely studied in psychology and in behavior sciences, as they are an important element of human nature. Nowadays they have also attracted the attention of researchers of computer science, especially in the field of human computer interactions. Advancement in textual analysis have allowed the area of emotion detection to become a recent interest in computational linguistic. Emotion detection is the newer area of textual analysis and therefore, has weaker standard methods. Emotion can be expressed as happiness, sadness, anger, disgust, fear, surprise and so forth. While board topic of emotion has been studied in psychology for decades[2], very little effort has been spent on attempting to detect emotion from text. In this work, we assume that emotion reaction of an input sentence is essentially represented by its word appearance.

II. RELATED WORK

The concept of affective computing in 1997 by Since Picard proposed that the role of emotion in human computer interaction. This domain attracted many researchers from computer science, biotechnology, psychology, and cognitive science and so on[3][5]. This sub-section outlines some lexical resources that researchers have compiled over the years to support affective computing and a verity of recently proposed methodologies. Lexical resources: One of the first such resources was a list of 1,336 adjectives manually labeled in "Predicting the semantic orientation of adjectives" in the year 1997. Word Net-Affect was introduced hierarchy of affective domain labels in "Word Net-affect: an affective extension of word net" in the year 2004. The subjectivity lexicon developed by is comprised of 8,000 words. Motivated by the assumption that different senses of the same term may have different opinion-related properties developed Sentimental Word Net, a lexicon based on word net in the year 2006. An automatically generated lexicon called Sentiment Full database was introduced in "Sentiment Full" Generating a reliable lexicon for sentinel analysis" in the year 2009.

a) Emotion Detection Approaches: Emotion detection approaches can be broadly classified into keyword-based, linguistic rules-based and machine learning technique.
b) Keyword-based Approaches using Affect Lexicon: Keyword based approaches are applied at the basic word level. Such a simple model cannot cope with cases where affect is expressed by interrelated words.
c) Linguistic Rules-based Approaches: The ESNA system was developed to classify emotions in news headlines. Manually added seed words to emotion lists and created a few rules in their system UPAR7, which identifies what is being said about the main subject and boosts its emotion rating by exploiting dependency graph.
d) Machine Learning Approaches: To overcome the limitations faced by rule-based methods, researches devised some statistical machine learning technique which can be sub-divided into supervised and unsupervised techniques.
e) Supervised machine learning with affect lexicons: One of the earliest supervised machine learning methods, where they used a hierarchical sequential model along with

Sentimental Word Net list for fine-grained emotion classification [5]. Blog sentences have been classified using Support Vector Machine (SVM)

f) Unsupervised machine learning with affect lexicons: An evaluation of two unsupervised techniques using Word Net-Affect exploited a vector space model and a number of dimensionality reduction methods. News headlines have been classified using simple heuristics and more refined algorithms.

III DIFFICULTIES

Many current approaches to emotion detection are based on supervised learning methods, in which large set of annotated data (where text has been labeled with emotions) is needed to train the model. Although the supervised learning methods can achieve good results, the availability of large annotated data set is very low and a model trained on one domain does not translate well to another. Some methods do not use supervised learning, but most of these methods use manually designed dictionaries of emotion keywords. A problem with such an affect lexicon-based method is that the number of emotion categories is fixed and limited in the dictionary. Another problem is that if a sentence expresses emotion using words that do not appear in the dictionary, then it would be considered to be unemotional. There are also methods that rely on linguistic rules, but designing such rules is not a trivial task. In addition, most of the current emotion detection methods look at individual words without considering the context a word is in. However, a word can invoke different emotions in different context. The work is organized as following: Language Processing entitled „Language Processing" describes about the Natural Language Processing (NLP) and the basics of different type of English sentences [7][2]. Emotion Estimation titled „Methodologies" describes about different emotion detection's approaches. Methodologies entitled „Emotion Estimation" Experiment entitled „Experiment" Result & Discussion entitled „Result & Discussion" Conclusion entitled „Conclusion" Future Work entitled „Future Plan" describes the future plan of our project

IV LANGUAGE PROCESSING

Natural Language Processing: Natural language processing (NLP) is the computerized approach to analyze text that is based on both a set of theories and a set of technologies. It is concerned with the interactions between computers and human (natural) languages. NLP is presenting naturally occurring texts at one or more level of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications[6].

NLP has two major methods of analysis:

1. Keyword Analysis or Pattern matching technique.
2. Syntactic driven parsing technique.

Keyword Analysis: In keyword analysis or Pattern matching technique, the system scans the input sentences for "selective" keywords and once they are encountered, the system responds with a „"built-in" reply.

V IMPLEMENTATION DETAILS

Preprocessing of Documents:

1. Lexical Analysis
2. Stop Word Removal
3. Collecting Nouns and Verbs
4. Stemming
5. Corpus Generation
6. Synonyms Root Density Representation
7. Pruning
8. Weighting

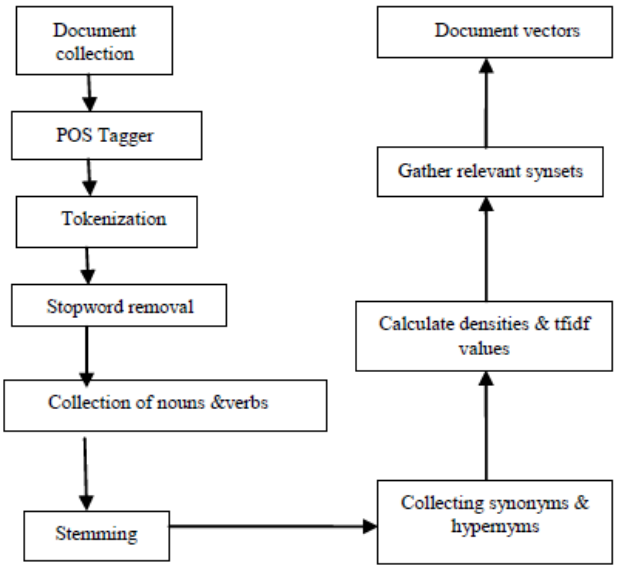

Fig: Preprocessing of Documents

VI. PROPOSED SYSTEM AND METHODOLOGY

The proposed system has 2 stages:

The first stage has to preprocess the documents, i.e. converting the documents into appropriate needs of data schemes. The second stage has to analyze the available data from first stage and divide it into clusters. This process is carried out by clustering algorithm [2].

Stage1. Preprocessing of the input documents :

The preprocessing consists of steps shown in figure:

1. The POS Tagger
2. Stop Words Removal
3. Stemming
4. Collecting Synonyms & Root Word
5. Calculate Frequency (Weighting)

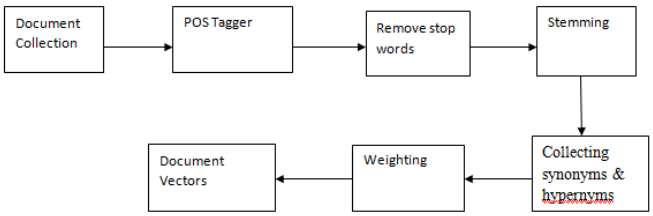

Fig: Process Flow

In this document set is taken as input and converted into form suitable for processing.

Each document must be transformed into a feature vector which can be presented to a machine-learning system. Adding semantic information to each document using external sources like word net gives better clustering results. Using this Word Net information, a better representation in the form of document vectors for documents can be obtained.

1. The POS tagger [7]: It relies on the text structure and morphological differences to determine the appropriate part-of-speech. A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc.

2. Stop Words Removal: To save space and to speed up searching process, the words which are considered as less important should be removed. Any group of words can be chosen as stop words such as 'the',' at' ,'on',' which'[5].

3. Stemming: Words with the same meaning appear in various morphological forms. Stemming algorithm is used to reduce the word to its root or stem. The key terms used in document are expressed by stem rather than original words. For example, consider the words "playing", "played", "play", and "player" can be reduced to the root word, "play"[5][7].

4. Collecting synonyms & Root Word: The stemmed words are looked up in the word net [8] and their corresponding synonyms and root words are added. Infrequently occurring synonyms sets are discarded, and those that remain form the feature set.
5. Weighting[9]: Weights are assigned to give an indication of the importance of a word on each document.

VIII. RESULTS

For Performance evaluation of the approach we measure it based on 2 parameters, precision and recall. Precision and Recall are defined in terms of a set of retrieved documents (e.g. the list of documents produced for a query) and a set of relevant documents (e.g. the list of all documents that are relevant for a certain topic)[1][5][7]

1. PRECISION:

Precision is the fraction of retrieved documents that are relevant to the find.

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Fig. Precision Graph

2. Recall:

Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

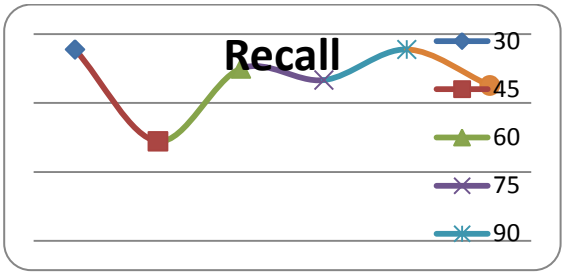

Fig. Recall Graph

VII. CONCLUSION AND FUTURE WORK

Emotion Detection can be seen as an important field of research in human-computer interaction.A sufficient amount of work has been done by researchers to detect emotion from facial and audio information whereas recognizing emotions from textual data is still a fresh and hot research area.In this paper, methods which are currently being used to detect emotion from text are reviewed along with their limitations and new system architecture is proposed, which would perform efficiently.

REFERENCES

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis , S. Kollias, "Emotion recognition inhuman-computer interaction," in IEEE Signal Processing Magazine, vol. 18(1), Jan. 2001, pp. 32-80,doi: 10.1109/79.911197

[2] Parrott, W.G, "Emotions in Social Psychology," in Psychology Press, Philadelphia 2001

[3] C. Maaoui, A. Pruski and F.Abdat, "Emotion recognition for human machine communication", Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 08), IEEE Computer Society, Sep. 2008, 1210-1215,doi: 10.1109/IROS.2008.4650870

[4] C. Yang, K. H.-Y. Lin and H.-H. Chen, "Emotion classification using web blog corpora," Proc. IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society, Nov. 2007, pp. 275-278, doi: 10.1109/WI.2007.50.

[5] C. M. Lee, S. S. Narayanan, and R. Pieraccini, "Combining Acoustic and Language Information for Emotion Recognition," Proc. 7th International Conference on Spoken Language Processing (ICSLP02), 2002, pp.873-876.

[6] C.-H. Wu, Z.-J. Chuang and Y.-C. Lin, "Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models," ACM Transactions on Asian Language Information Processing (TALIP), vol. 5, issue 2, Jun. 2006, pp. 165-183, doi:10.1145/1165255.1165259.

[7] C. Elliott, "The affective reasoner: a process model of emotions in a multiagent system," in Doctoral thesis, Northwestern University, Evanston, IL, May 1992

[8] Protégé tool, www.protege.stanford.edu/,

[9] Nicu Sebea, Ira Cohenb, Theo Geversa, and Thomas S. Huangc "Multimodal Approaches for Emotion Recognition: A Survey", USA

[10] http://sail.usc.edu/~kazemzad/emotion_in_text_cgi/DAL_app/index.php?overall=bad&submit_evaluation=Submit+Query".