Sentiment Computing for Visual Emotion Generation on Social Media using Text Mining

Nayana More[1], Deepali Jadhav[2]

[1]*P.G. Student, Department of Computer Science & Engineering, KIT's College of Engineering, Maharashtra, India*[1]

[2]*Assistant Professor, Computer Science & Engineering, KIT's College of Engineering, Maharashtra, India*[2]

Abstract

The fast increase of the World Wide Web has helped increased online communication and opened up newer streets for the general public to post their opinions online. This has led to a generation of large amounts of online content rich in user opinions, sentiments, emotions, and evaluations. We need computational approaches to successfully analyse this online content, recognize and aggregate relevant information, and draw useful conclusions. Much of the current work in this direction has typically focused on recognizing the polarity of sentiment (positive/negative). In this writing, we have suggested a system that recognizes the emotion from the text of Social networking websites by using a modified approach that uses affective word based and sentence context level emotion classification method. Also to adequately express the emotion of a user we have developed a visual image generation approach that generates images according to emotion in text.

Keywords - *Emotion, Sentiment, Classification, Social Networking, Recognizing, Accuracy, Extraction, Tagging, Detection, Pre-processing;*

I. INTRODUCTION

Emotion recognition is currently widely being studied, be it detection from facial expressions, from textual information, or speech. The most common form of communication on the web is in the form of text, offering a platform for computer systems to behave more intelligently based on the user's mood. With large amounts of textual data available in the form of blogs, emails, etc., a better human-computer interaction system needs to be able to analyse the text and infer the sentiment/emotion of the user. Although communication systems can identify the user's emotional states from different communication modalities, the variety and complexity of language make it difficult for researchers to recognize emotional states from pure textual data.

The application areas of textual emotion detection are manifold:

A. Sentiment Analysis

Sentiment Analysis is a widely pursued research area today, with companies valuing consumer opinions about their products. Sentiment Analysis aims at inferring the sentiment (positive or negative) of the consumer based on his/her review.

B. Text to Speech Generation

Human-Computer interactions systems aim to communicate more humanely through the synthesis of speech from text. To make the speech lifelike, the emotion behind the text has to be inferred. This makes Text to Speech Generation a very fruitful area of research.

C. Better Computer Interaction System

Many kinds of the communication systems, such as dialogue systems, automatic answering systems and human-like robots, can apply emotion recognition techniques so that a user feels as if the system is more human. A better response system, based on the user's current mood/emotion, makes users and computers work in sync.

Our work aims at recognizing the emotions from the text of Social Networking websites, i.e. Blog and tweet data. The sentence level emotion classification proposed in this paper is a tougher problem than document-level emotion classification. Documents contain a more significant number of words, and a keyword-based approach has a better chance to capture the emotion due to the presence of larger number keyword instances. A sentence, on the other hand, has very few keyword instances.

The novel approach suggested in this paper deals with finding out the emotion by using affective word based emotion classification and by using the sentence context analyser for emotion classification. To capture the semantics of the sentences, such as subject-verb and object-verb instances it is needed to analyse the sentence meaning. Also to make the effective Human-computer interaction or to adequately express the feeling of user animated agents are generated based on the emotions of textual Interactions

II. RELATED WORK

Much work has been done in the area of emotion classification with the majority being Lexicon-based Supervised Learning approaches.

Strapparava et al. (2) provide an unsupervised approach to emotion detection, experimenting with several knowledge-based and corpus-based methods, including Latent Semantic Analysis (LSA) used in conjunction with Word Net. Corpus knowledge is incorporated from annotated blog data.

M. Ishizuka [16] presents a novel approach to Emotion Estimation that assesses the affective content from textual messages. The primary purposes of this work are to detect emotion from chat or other dialogue messages. An advanced keyword spotting technique recognizes the affective content of the textual message.

C. O. Alm, D. Roth, and R. Sproat [1] explore the text-based emotion prediction difficulty empirically, using supervised machine learning with the SNoW learning architecture. They have proposed a paragraph-level emotion classifier by feeding a machine learning algorithm with training corpus. The purpose of this article is to analyse the emotional relationship of sentences in the old domain of children's fairy tales for the following usage in an appropriate expressive rendering of text-to-speech synthesis.

S. Aman and S.Szpakowicz [19] proposes an approach to examine the expression of emotion in language through a corpus annotation study and to prepare an annotated corpus for use in automatic emotion analysis experiments. They also explore computational techniques for emotion classification. First, they prepared a list of seed words for six basic emotion categories proposed by Ekman, using the seed words for each category and retrieved blog posts containing one or more of those words. This approach takes into account some features such as the affective words, punctuations, the theme of the story, and storylines, etc. The first results appeared to be promising when given a sufficient number of input sentences of a paragraph within the narrative domain of children's fairy tales. However, the applicability of such emotion sensing technique to domain-independent texts in sentence-level would not be possible.

Mohamed Yassine Hazem Hajj [9] proposes a new framework for characterizing emotional interactions in social networks and then uses these characteristics to distinguish friends from acquaintances. The interest of this paper is to find out whether the text is an emotional expression or of writers emotion or not.

Shenghua Bao, Shengliang Xu, Li [3] presents and analyse a new problem called social, affective text mining, which aims to identify and model the connections between online documents and user-generated social emotions. They have proposed a joint emotion-topic model by augmenting Latent Dirichlet Allocation with an intermediate layer for emotion modeling. The emotion-topic model proposed in this paper allows associating the terms and emotions via topics which is more flexible.

Wu et al. [15] proposed a novel approach for sentence level emotion detection based on the semantic labels (SLs) and attributes (ATTs) of entities of a sentence. To distinguish the emotions of happy and unhappy, the SLs are manually classified into three categories, Active SLs (e.g., obtain, reach, lost, hinder), Negative SLs (e.g., no, never), and Transitive SLs (e.g., finally, but, fortunately). ATTs of an entity is obtained automatically from a lexical resource, Word Net (Fellbaum [6]). The results show the degree of accuracy is rather high. The proposed approach exploited modern Natural Language Processing (NLP) technologies and is one of the rare studies that dealt with sentence-level emotion detections with high precision. However, the approach required the use of affectively annotated corpus which is not readily available in a broad context. Furthermore, only two emotions, happy and unhappy have been achieved

Lu et al. [7] recently proposed an approach that detects event-level text by analyzing the mutual action histograms of event entities. The approach is divided between training and a testing phase to detect positive, negative, and neutral emotions. In the training phase, some reference entity pairs are identified manually, and verb categorization processes are implemented manually to form a set of verb synonym groups. To handle the events associated with each of the reference entity pairs and verb synonym groups, a set of emotion generation rules (EGRs) are constructed manually for all the verb synonym groups. In the testing phase, to form the dataset to assess the robustness of the approach, several lexico-syntactic patterns are employed to collect a large number of entities from web pages. Lu's approach achieved a high precision rate of approximately 75% to detect positive, negative, and neutral emotions.

Hu *et al.* , performed a Bayes text classification.
Their results show that the Naive Bayes classification method achieves high performance on text classification. However, this method categorizes texts into only two classes: positive and negative, which excludes the reader's emotions such as angry, happy etc.

Yue Ning, Yan Wang In this paper, we propose a $\chi 2$-based Chinese text emotion classification by using Naive Bayes five sentiment categories. We run two experiments; one uses sentiment words extracted from HowNet and Chinese thesaurus: TongYiCi CiLin,

III. PROPOSED WORK

The overall framework of the proposed system consists of Pre-processing of data, finding out the emotions from powerful words of sentences, if the sentence does not contain any useful word, then emotion recognition will be done by analysing the sentence structure for finding the relation between verb, subject, an object.

A. Data Collection

This dissertation work intends to be able to recognize emotions from text automatically. This requires an appropriate corpus of text that can be used for training and testing in emotion recognition experiments. For training machine learning systems and the evaluation of any automatic learning system, it is pre-requisite to have annotated data. Blogs and tweets are online personal journals containing owner's reflections and comments. They make a good candidate for emotion study, as they are likely to be rich in emotional content. We now describe the collection we used for our experiments.

We have used the datasets with sentence-level annotations of emotions include about 400 sentences from blogs, compiled by Aman and Szpakowicz (2007); The blog data they have collected from the Web in the following manner.

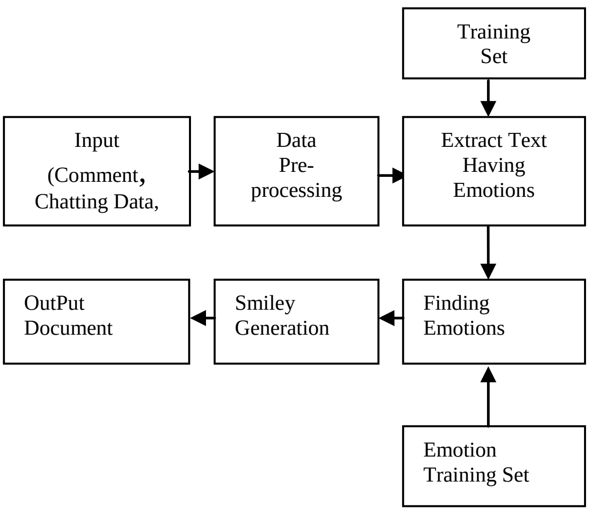


Fig 1 . Framework of system

First, a set of seed words was identified for each of the emotion categories. In preparing the set, I took words that are commonly used in the context of a particular emotion. Thus, I chose words such as “happy”, “enjoy”, “pleased” as seed words for the *happiness* category; “afraid”, “scared”, “panic” for the *fear* category, and so on. Next, the seed words for each category were fed to the blog search engine, BlogPulse16 and blog posts including one or more of those words were retrieved. A total of 173 blog posts were collected in this manner.

Also, we have used The Archivist5 a free online service that helps users extract tweets using Twitter's Search API.6 For any given query, Archivist first obtains up to 1500 tweets from the previous seven days. Subsequently, it polls the Twitter Search API every few hours to obtain newer tweets that match the query. We supplied Archivist with the different emotions and collected about 300 tweets. We discarded tweets that had fewer than three valid English words. For this collected data the annotation process is used for attaching the particular emotion with the sentence. The emotion annotation is essential because the emotionally annotated sentences are then further used for training of a machine learning algorithm.

B. Pre-processing of Data

Before predicting with our system, raw texts should be pre-processed. A specific procedure is as follows. First, all web documents are pre-processed. First, the Training article is segmented into terms.
Output words are filtered through a manual collected Stop Words Dictionary. To get out the emotion from a text all additional content must be eliminated so it is needed to remove the stop words that bear no meaning about Emotion and the text put into an array. The words irrelative with emotion such as "an , the, and, before,..." are removed from the Sentence.

The words obtained from above processing of the data is found to be too sparse to be useful as features for classification as they do not generalize well. The typical reason for this is the presence of a large number of inflections of the same word. Hence, the root form of a word is to be extracted as a feature. The conventional technique used to find the root form of a word is Stemming. Stemming is the method of reducing inflected words to their stems. For eg., stemmer, stemmed and stemming are all reduced to their root word 'stem.' The stem need not be equal to the morphological root of the word. For eg., cookery is stemmed to cookery, which isn't even an actual word. Also, stemming doesn't do dictionary lookups to identify the actual root form of the word but instead strip the ending characters to produce a stemmed form of the word. For eg., better is not stemmed to good. stemming is computationally less expensive as well as has greater coverage as it doesn't look at the contextual meaning of words and works without part of speech tagging.

The algorithm used for stemming is Porter stemmer. The stemmer methods are classified into customs where each of these customs deals with a specific suffix and having a satisfied condition(s) to meet.

EED ->EE -> agreed->agree

(*v*) ING -> motoring ->motor

After the pre-processing, each article is represented as a set of meaningful terms.

1. APPROACH 1: Detection of Emotional features using affective words Lexicon: Classify the sentence using naïve Bayes

The system divides a text into words and delivers an emotional estimation for each of these words as in the following figure:

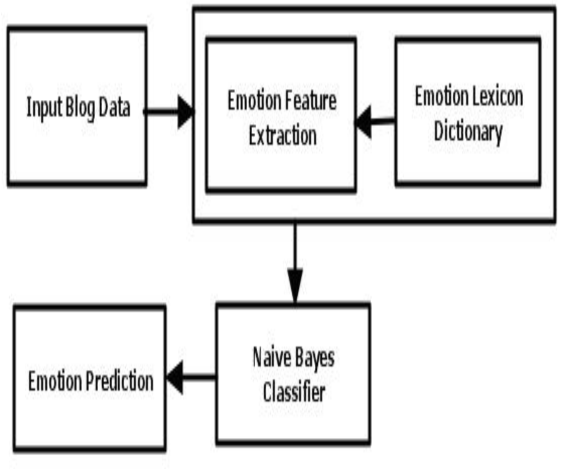


Figure 2: Basic block Diagram Emotion Classification

The extraction of the text having emotion is the most critical step in emotion recognition. The extracted text from the sentence which represents the emotion of that sentence is then further used for training of a machine learning system and finding out the emotion from given text.

In the step of extracting the text having emotion, it needs to find out the feature set from a particular sentence. This feature set is nothing but those features, which distinctly characterize emotional expressions, Emotion words are taken as a feature set in [4]. Here, we also use emotion words and build an emotion dictionary for every emotion class. Each example or sentence has one or more properties, which are called features. These features describe the properties of the examples and can be used in learning as predictors of the target class.

The features that represent the emotion of text are derived from an emotion lexicon dictionary. This emotion lexicon dictionary contains the words that are accounting for the emotion in the sentences. We have prepared an emotion lexicon dictionary in a semi-automatic way by using Word Net-Affect Database with Word Net 1.6 [4] and SentiWordNet Dictionary. To prepare the Affective word Lexicon Dictionary, we first find synonyms sets of affective words. We select the emotion words according to emotion categories, i.e., "happy", "sad" , "angry" , "fear," then use keywords recognition and synonyms to choose related words from Word Net based on the affective words set.

Category	**Sample Words**	**Count**
Happy	joy, love, rejoicing, happiness, cheerful, happy, enjoy, good, nice, excited	**207**
Sad	sorrow, misery, heartbreak, death, cry, unhappy, depress, bad, miss, wept	**190**
Anger	angry, annoyed, pissed, mad, livid, displeasingly, aggressive, hateful, hostile	**97**
Fear	scary, fright, terror, frightful, terrible, intimidate, dread, afraid, dangerous	**95**

Table 1: emotional words

The words in sentences are checked in the emotion lexicon dictionary, the words that are present in the lexicon dictionary forms the feature set for that sentence.

To extract the text having emotion the following process is done:

Sentence	**Emotional words**	**Emotion**
Today I am very happy!	Happy	Happy
It was very annoying Thing	Annoying	Anger
I was so scared at that moment.	Scared	Fear

Table 2: sentence and its emotional words

Let the sentence S is composed of n terms {t1, t2, ..., tn.} A feature set FS for particular sentence S is defined as follows:

FS={tk | tk, 1<=k,=n, tk is part of Lex}

For those tk (1<=k<=n) appearing in the lexicon dictionary

After finding out a set of features from a sentence, this feature set is then mapped to an emotion category. Then an emotion assignment process is transformed into a classification problem using Naïve Bayes Classifier. Naive Bayesian [8] calculates the probability of a word appeared in a category using Naïve Bayes Classifier.

2. APPROACH 2: Detection of Emotional features using Sentence context level:

The word spotting system is too simple to deal with sentences such as

"I think that he is happy" or "the lake near my city was beautiful,

Since here, the mouthpiece is not necessarily happy, or the speaker's emotion depends on the semantics of the sentence deeply. For example, if the sentence is
I saw a big LION,
This sentence is not containing any emotional word as the previous sentences, so to accurately find the emotion of such sentence it is needed to consider the context structure of a sentence. We hence perform the latter steps on *sentence-level processing.*

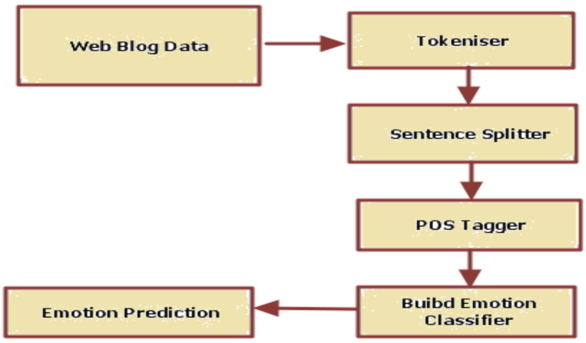


Figure 3: sentence-level processing for emotions In this step, the multiple-sentence text is spited into single sentences. Each sentence is estimating the emotion separately.

3. POS (part of speech) Tagging and Sentence Arrangement Recognition

In this step, the syntactic expression types are determined out from parse trees that are created by the parser. The phrase types express the semantic positions in the sentence, for examples, a noun phrase (NP), verb phrase (VP). In the top-level of the decision, the syntactic subject and verb can be derived from the noun phrase and verb phrase.

In a typical English sentence, the verb's syntactic arguments are usually associated with the participants of the event. A semantic role is the relationship of a syntactic argument with the verb. In our approach, when the emotions for free-text sentences are to be detected automatically, the semantic roles, at least including the subject, verb, an object of an event need to be obtained. Each sentence wrote by the user is parsed by to retrieve the verb and objects of the events. These verb and objects are then used for the semantic analysis of the sentence. The system then allocates the emotion based on the verb and the affection categories of the objects which are classified to different useful categories.

4. Emotion Classification

The Naive Bayes classifier is well studied and is extensively used because it is fast, easy to implement and relatively effective, even though the far-reaching independence assumptions are often inaccurate. We have applied the Naïve Bayes classification of them.
This method decides that class (in our case: that speech style) which has the highest conditional probability provided given features.

It makes each word *wi* of document *d* in two sampling levels, i.e., sample an emotion *ei* according to the emotion cycle count *γd*, and sample a word *wi* given the emotion under the conditional probability
$P(w|e)$.

$$P(w|e) = \frac{|(w,e)|}{\sum_{w' \in W} |(w',e)|},$$

The model parameters can be learned by maximum probability estimation. In particular, the conditional probability of a word *w* given an emotion *e* can be estimated as follows,

$$|(w,e)| = S + \sum_{d \in D} \delta_{d,w} \cdot \gamma_{d,e},$$

Where $|(w, e)|$ is the co-occurrence count between word $w \in W$ and emotion $e \in E$ for all the documents. It
can be formally obtained based on the word and emotion frequency counts, and for predicting emotion on a new document *d*, we can apply the Bayes theorem under the term independence assumption

$$\text{P (e|d)} = \frac{P(d|e)P(e)}{P(d)} \alpha \ P(d|e)P(e)$$
$$= P(e) \prod_{w \in d} P(w|d)^{\delta_{d,w}}$$

Where $P(e)$ is the a priori probability of emotion *e*. It can again be calculated by maximum likelihood estimation (MLE) from the emotion distribution of the entire collection.

$$\hat{H} = \arg\ \max p(H|X)$$

Using the Bayesian rule, it can be written as

$$\hat{H} = \arg\ max \frac{p(X|H)p(H)}{p(X)}$$

Using the fact that LR images *X* are known (their probability is constant), and they are independent of each other, the above equation becomes

$$\hat{H} = \arg\max \prod_{i=1}^{m_3} p(X_i|H)p(H)$$

The following table shows the Mean classification accuracy shared between all emotions considered in thiswork. This reveals the accuracy with a comparison of the proposed algorithm results of the other references mentioned

C. Visual Emotion Generation from the text

The last step in Emotion recognition is visual Image generation related to a particular emotion. This is an effective way of communicating with followers and friends because sometimes words are not enough to clearly express the feeling.

1. Result and Analysis

The proposed approach result for the for each emotion
Category is as follows:

Emotion	Precision	Recall	F-Measure
Happy	89.55	95.40	92.38
Sad	96.60	88.33	92.28
Anger	89.20	95.83	92.39
Fear	88.70	93.33	90.93

Table 3: Precision-Recall and F-Measure

The Emotion classification module analyses the emotion tendency of the incoming messages and returns the result to the Visual Image Generation Module. By applying the above emotion estimation method on textual data of Social Networking Websites, the multiple images are retrieved according to emotion and words of the sentence then the system can predict the particular image.

Category	Accuracy
Happy	92.38
Sad	92.28
Anger	92.39
Fear	90.93
Average	91.85

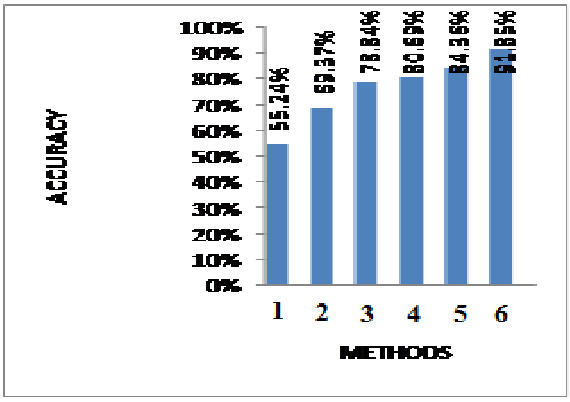
Table 4: Accuracy of system

The given table and Graph shows the classification accuracy with Personal emotion and different methods in table 4 and 5 respectively.

The proposed system uses much accuracy over the earlier method. Here proposed system having 91.85 % accuracy.

Method ID	Method	Accuracy
1	Hierarchical Classification BoW+SO	55.24%
2	All features + sequencing(same-tune-eval)[9]	69.37%
3	Unigram with Presence	78.84
4	Simple Linguistic Processing using SVM[11]	80.69
5	Customized decision tree algorithm[13]	84.36%
6	Proposed Approach	91.85%

Table 5: Accuracy with comparison of different Methods



Graph 1: Accuracy with comparison of different methods

IV. CONCLUSION

This rapid growth of the World Wide Web led to the generation of large amounts of online content rich in user opinions, sentiments, emotions, and evaluations. Here computational approaches applied successfully and analysed this online content, recognize and aggregate relevant information, and draw useful conclusions. Much of the current work in this direction has typically focused on recognizing the polarity of sentiment (*positive*/*negative*). In this paper, we have introduced a system that recognizes the emotion from the text of Social networking websites by using a modified approach that uses the operative word based and sentence context level emotion classification method. Also to adequately express the emotion of a user we have developed a visual image generation approach that generates images according to emotion in text.

REFERENCES

[1] C. O. Alm, D. Roth, and R. Sproat, "Emotions from the text: machine learning for text-based emotion prediction," in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Vancouver, British Columbia, Canada: Association for Computational Linguistics, 2005, pp. 579–586.

[2] C. Strapparava and R. Mihalcea, "Learning to identify emotions in text," in Proc. of the 23rd Annual ACM Symposium on Applied Computing, SAC'08, 2008, pp. 1556–1560.

[3] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu "Mining Social Emotions from Affective Text".

[4] Cheng-Zhong Xu, Tamer I. Ibrahim, "A Keyword-Based Semantic Prefetching Approach in Internet News Services" IEEE conference on Knowledge and Data Engineering, vol. 16, no. 5, May 2004

[5] C. Yang, K. H.-Y. Lin, and H.-H. Chen, "Emotion classification using web blog corpora," in Proc. of the 2007 IEEE / WIC / ACM International Conference on Web Intelligence, WI'07, 2007, pp. 275– 278.

[6] A. J. Gill, D. Gergle, R. M. French, and J. Oberlander, "Emotion rating from short blog texts," in Proc. of CHI'08 on Human factors in computing systems, 2008, pp. 1121–1124.

[7] Cheng-Yu Lu , William W.Y. Hsu, Hsing-Tsung Peng, Jen-Ming Chung " Emotion Sensing for Internet Chatting: A Web

Mining Approach for Affective Categorization of Events" 13th IEEE International Conference on Computational Science and Engineering, 2009

[8] C. Strapparava and A. Valitutti, "Wordnet-affect: an affective extension of wordnet," in Proc. of the 4th international conference on Language Resources and Evaluation, LREC'04, 2004

[9] Mohamed Yassine, Hazem Hajj, "A Framework for Emotion Mining from Text in Online Social Networks", IEEE International Conference on Data Mining Workshops, 2010.

[10] R. Cai, C. Zhang, C. Wang, L. Zhang, and W.-Y.Ma., "Music sense: contextual music recommendation using emotional allocation," in Proc. of the 15th international conference on Multimedia, 2007, pp. 553–556.

[11] Amitava Das, Sivaji Bandyopadhyay, "Theme Detection an Exploration of Opinion Subjectivity", IEEE International Conference on Data Mining Workshops, 2009.

[12] DR. Yashpal Singh, Alok Singh Chauhan, "neural networks in data mining" Journal of Theoretical and Applied Information Technology,vol.5, pp.37- 42, 2005 – 2009

[13] Lina L. Dhande1 and Dr. Prof. Girish K. Patnaik2, "Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier",International Journal of Emerging Trends & Technology in Computer Science (IJETTCS). Volume 3, Issue 4 July-August 2014 ISSN 2278-6856.

[14] Mukesh C. Jain, V. Y. Kulkarni, "TexEmo: Conveying Emotion from Text- The Study", International Journal of Computer Applications (0975 – 8887) pp.43-50, Vol 86 – No 4, January 2014.

[15] Wu, Chung-Hsien & Chuang, Ze-Jing & Lin, Yu-Chung. (2006). Emotion recognition from text using semantic labels and separable mixture models. ACM Trans. Asian Lang. Inf. Process.. 5. 165-183. 10.1145/1165255.1165259.

[16] Ma, Chunling & Prendinger, Helmut & Ishizuka, Mitsuru. (2005). Emotion Estimation and Reasoning Based on Affective Textual Interaction. 622-628. 10.1007/11573548_80.

[17] Yassine, Mohamed & Hajj, Hazem. (2010). A Framework for Emotion Mining from Text in Online Social Networks. Proceedings - IEEE International Conference on Data Mining, ICDM. 1136-1142. 10.1109/ICDMW.2010.75.

[18] Saima Aman1 and Stan Szpakowicz "Identifying Expressions of Emotion in Text", School of Information Technology and Engineering, University of Ottawa, Ottawa,

[19] Rutuja Karkar1, Shweta Nagdev, Pranjal Gangrade, Deepali D. Gatade"Transformation of Sentimental Impact for Documents",International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072.