ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 6, Issue 2, February 2018**

# A Survey on Emotions Generation using Text Mining for Social Networking Websites

Nayana More [1], Deepali Jadhav [2]

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

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

**ABSTRACT**: The rapid growth of the World Wide Web has facilitated increased online communication and opened up newer avenues for the general public to post their opinions online. This has led to generation of large amounts of online content rich in user opinions, sentiments, emotions, and evaluations. We need computational approaches to successfully analyze 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 proposed a system that recognizes the emotion from text of social networking websites by using a modified approach that uses affective word based and sentence context level emotion classification method. Also to effectively express the emotion of a user we have developed visual image generation approach that generates images according to emotion in text.

.

**KEYWORDS**: Emotion, Sentiment, Classification, Social Networking, recognizing;

## I. INTRODUCTION

Emotion recognition is currently widely being studied, be it detection from facial expressions, from textual information, or from speech. The most prevalent form of communication on 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 sys- tem needs to be able to analyze the text and infer the sentiment/emotion of the user. Although communication systems can identify the users emotional states from different communication modalities, the variety and complexity of language makes it difficult for researchers to recognize emotional states from pure textual data.

The application areas of textual emotion detection are manifold:

1. **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.
2. **Text to Speech Generation**: Human- Computer interactions systems aim to communicate in a more humane way through the synthesis of speech from text. In order 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.
3. **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 feel 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 a Social Networking websites i.e Blog and tweet data. The sentence level emotion classification proposed in this paper is tougher problem than document-level emotion classification. Documents contain larger 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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 6, Issue 2, February 2018**

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

## II. RELATED WORK

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

In [1] presents a novel approach to Emotion Estimation that assesses the affective content from textual messages. The main goals of this work are to detect emotion from chat or other dialogue messages. The affective content of the textual message is recognized by an advanced keyword spotting technique.

In [2] explores the text-based emotion prediction problem 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 goal of this paper is to classify the emotional affinity of sentences in the narrative domain of children fairy tales for subsequent usage in appropriate expressive rendering of text-to-speech synthesis.

In [3] presents and analyze a new problem called social affective text mining, which aims to discover 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.

In [4] proposes an approach to investigate the expression of emotion in language through a corpus annotation study and to prepare (and place in the public domain) 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 a number of features such as the affective words, punctuations, theme of story, and story lines, etc. The primitive results appeared to be promising when given sufficiently 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.

In paper [5] 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 WordNet. Corpus knowledge is incorporated from annotated blog data.

In [6] 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. the paper presents a new perspective for studying friendship relations and emotions expression in online social networks where it deals with the nature of these sites and the nature of the language used.

In [7] proposes a web mining approach for classification of movie reviews, this study proposes a system that extracts review of data, of movie blogs. It introduces an architecture, implementation and evaluation of a web blog mining application called the BlogMiner, which extracts and classifies people's opinions or sentiments from the contents of weblog about movie reviews.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 6, Issue 2, February 2018**

In [8] try to figure out the sentiment categories those articles on the web may produce in readers. In this paper, they have propose to classify unlabeled articles into sentiment categories. After investigating several classic feature selection methods and classification algorithms, they use the emotion lexicon dictionary for feature selection, and train Naive Bayesian classifier. The text data they have collected was from the Chinese social news,

In [9] recently proposed an approach that detects event-level text by analyzing the mutual action histograms of an event entities. The approach is divided in a training and a testing phase to detect positive, negative, and neutral emotions. In the training phase, a number of 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.

In [10] 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 are obtained automatically from a lexical resource, WordNet (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 a use of affectively annotated corpus which is not readily available in a wide context. Furthermore, only two emotions, happy and unhappy have been achieved

## III. PROPOSED FRAMEWORK

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

### A. Data Collection

The objective of this dissertation work is to be able to automatically recognize emotions from text. 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 for the evaluation of any automatic learning system, it is pre-requisite to have an annotated data. Blogs and tweets are online personal journals containing owner's reflections and comments. They make good candidate for emotion study, as they are likely to be rich in emotion 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. 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 containing 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 is 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

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

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 important because the emotionally annotated sentences are then further used for training of a machine learning algorithm.

B. **Pre-processing of Data:**

Before predicating with our system, raw texts should be pro-processed. Specific procedure is as fellows. First, all web documents are pre-processed. First the Training article are segmented into terms.
Output words are filtered through a manual collected Stop Words Dictionary. To find out the emotion from a text all unnecessary content must be removed 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 common 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 feature. The common technique used to find the root form of a word is Stemming. Stemming is the process of reducing inflected words to their stems. For eg., stemmer, stemmed and stemming are all reduced to their root word 'stem'. The point to be noted is that the stem need not be identical to the morphological root of the word. For eg., cookery is stemmed to cookeri, which isn't even an actual word. Also, stemming doesn't do dictionary lookups to identify the actual root from of the word but in- stead just 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 operations are classified into rules where each of these rules deals with a specific suffix and having certain condition(s) to satisfy.

| | | |
|---|---|---|
| EED ->EE | -> | agreed->agree |
| (*v*) ING | -> | motoring ->motor |

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

**APPROACH 1:**
**Detection of Emotional features using affective words Lexicon:**
**Classify the sentence using naïve bayes**
The system divides a text into words and performs an emotional estimation for each of these words as in following figure 1 the extraction of the text having emotion is most important step in emotion recognition. The extracted text from the sentence which represent 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 is need to find out the feature set from 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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 6, Issue 2, February 2018**

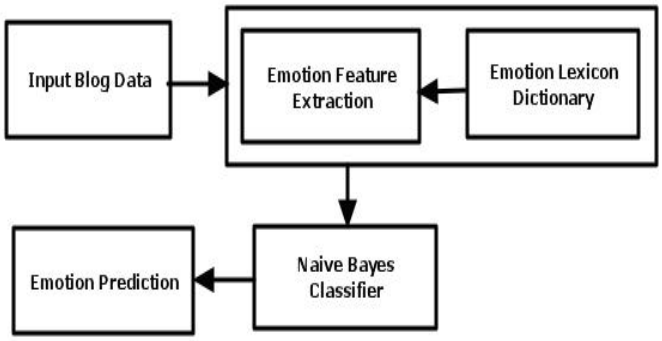


**Figure1: emotion prediction**

In Table 1 the features that represent the emotion of a 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 a emotion lexicon dictionary in a sem-automatic way by using WordNet-Affect Database with WordNet 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 key words recognition and synonyms to choose related words from WordNet 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** |

**Table1: emotional words category**

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.

In order to extract the text having emotion the following process is done:
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

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 6, Issue 2, February 2018**

After finding out set of features from a sentence, this feature set is then mapped to an emotion category. Then an emotion assignment process is transformed to a classification problem using Naïve Bayes Classifier. Naive Bayesian [8] calculates the probability of a word appeared in a category using Naïve Bayes Classifier. Table 2 used to predict actual emotion from the sentence or group of sentence input towards the system.

| 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: emotional words category prediction**

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

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

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

Since here, the speaker is not necessarily happy or the speaker's emotion depends on the semantic 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.

## IV. CONCLUSION AND FUTURE WORK

In this paper various methods for emotion generation and prediction studied and analysed. Also lead to generation of large amounts of online content rich in user opinions, sentiments, emotions, and evaluations. We finalise 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 paper we have proposed a system that recognizes the emotion from text of social networking websites by using a modified approach that uses affective word based and sentence context level emotion classification method. Also to effectively express the emotion of a user we have developed visual image generation approach that generates images according to emotion in text.

## REFERENCES

1. Chunling Ma,Helmut Prendinger,Mitsuru Ishizuka "Emotion Estimation and Reasoning Based on Affective Textual Interaction" International Conference on Affective Computing and Intelligent Interaction ACII: Affective Computing and Intelligent Interaction,pp.622-628, 2005.
2. C. O. Alm, D. Roth, and R. Sproat, " Emotions from text: machine learning for text-based emotion prediction" ,Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Vancouver, British Columbia, Canada: Association for Computational Linguistics,pp.579-586, 2005.
3. Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han,and Yong Yu ,"Mining Social Emotions from Affective Text" , IEEE conference on Knowledge and Data Engineering ,pp.1658-1670,2010.
4. Amine Trabelsi, Claude Frasson," The Emotional Machine a Machine Learning Approach to Prediction of User's Emotion and Intensity", ICML '02: Proc. 19th Int'l Conf. Machine Learning, pp. 315-322, 2002.
5. C. Strapparava and R. Mihalcea, "Learning toidentify emotions in text," in Proc. of the 23rd Annual ACM Symposium on Applied Computing, SAC'08, pp. 1556-1560.2008,42
6. Mohamed Yassine Hazem Hajj, "A Framework for Emotion Mining from Text in Online Social Networks", IEEE International

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 6, Issue 2, February 2018**

Conference on Data Mining Workshops, pp.1136-1142,2010,75.

7. Arzu Baloglu, Mehmet S. Aktas , "BlogMiner:Web Blog Mining Application for Classification of Movie Reviews", Fifth International Conference on Internet and Web Applications and Services ISBN 978-0-7695-4022-1,2010.
8. Yue Ning, Tingshao Zhu1, Yan Wang , "Affective-Word based Chinese Text Sentiment Classification" Graduate University of Chinese Academy of Sciences Beijing 100190, China. ISBN 978-1-4244-9142-1.
9. Cheng-Yu Lu,William W. Y. Hsu,Jan-Ming Ho,"Event-Level Textual Emotion Sensing Based on Common Action Distributions between Event Participants" International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems IEA/AIE: Advanced Research in Applied Artificial Intelligence, pp.427-436,2012.
10. 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,pp. 601 – 611,2004.
11. 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, pp. 275- 278,2007.
12. Yang shen,Shuchen LI,Ling zhengg,"Emotion Mining Research on Microblog",IEEE, Web Society, SWS '09. 1st IEEE Symposium on 2009 ISBN: 978-1-4244-4157-0,2009.
13. Sivaji Bandyopadhyay, "Theme Detection an Exploration of Opinion Subjectivity" ,Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on IEEE SBIN: 2156-8111,2009.
14. C. Strapparava and A. Valitutti, "Wordnet-affect: an affective extension of wordnet",in Proc. of th4th international conference on Language Resources and Evaluation, LREC'04,pp.1083-1086, 2004,.
15. DR. Yashpal Singh, Alok Singh Chauhan, " Neural networks in data mining",Journal of Theoretical and Applied Information Technology, ISSN 2277-1956/V1N3,pp.1449-1453,2009.