

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/303540705>

# Emotion Detection through Text

Conference Paper · April 2016

CITATIONS

0

READS

2,333

2 authors:



[Kashif Khan](#)

University of Camerino

6 PUBLICATIONS 3 CITATIONS

[SEE PROFILE](#)



[Muhammad Ejaz](#)

Minnesota State University Moorhead

20 PUBLICATIONS 73 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



IoT Big Data Analysis [View project](#)

## **Emotion Detection through Text: Survey**

**Kashif khan**

University of Bradford UK, University of Camerino Italy

&

**Sher Hayat**

University of Camerino, Italy

&

**Muhammad Ejaz khan**

University of Camerino, Italy

### **Abstract:**

*The survey will present recent work in the field of Emotion Detection through text. As to know that a lot of work is expected to be done in the field of Emotion detection through text. Emotion can be expressed through different ways it could be in the form of facial expression, speech or written text. Human can recognized the emotions through different ways but this is how a system recognized emotions. In this paper we will describe some recognition rules, methodologies like keyword based, lexical based, and knowledge based methods. We will apply some normalization techniques on sentences to know the emotion in a sentence. This survey report will present some limitations on text based emotions. This survey focuses on to solve somehow the problems which we are facing in emotion detection. Emotion plays an important role in social interaction and has strong connection with human body and brain signals. This survey will focus on SVM, KNN, HMM, Byes classification and will identify advantages and disadvantages.*

**Introduction:** Language plays an import role in the field of communication, as through languages you can express your feeling emotions etc. Emotion involves in feelings, behaviors, experience and cognitions. An emotion could be any strong feelings through some circumstances or mood or relationship. Exchange of emotion can be done through text, feelings, speech. Human can recognize their feelings, emotions but this is a challenge that how a computer recognize humans feelings in the form of text. Now here Human computer interaction plays an important role in the field of digitalization. Every second we have massive amount of data on internet now the challenge is how to digitalize this data and how to pick emotions. Emotion could be one word or bunch of words, it has no specific hierarchy to express emotion there is a book named “Emotion is Social physiology” where it has six Ekman [1] emotions which are joy, happiness, anger, surprise, fear, love. It has some other emotions which fall in secondary and territory category. We will work on sentence level to find emotions from text. While reading a book or sentence you can find the writer feeling through text. If “A person is happy” you can use positive feel positive response. If “the person is not happy” there would be different concept either the person is frustrated, sad or angry. So from text you can understand the user feeling. There are different applications like if an employee sends a harsh email to colleague or another employee when the employee read the email he will understand the email and through this way he will protect his job. Emotion detection can also help the marketer to help strategies for customer to build good relationship management, product service and product delivery. Psychologists can also get benefit from being able to infer people’s emotions based on the text that they write which they can use to predict their state of mind.

Recognition is a big challenge for human as well as for computer. Here sometimes you cannot recognize a person through their own emotions. While machine need some advanced algorithm to recognize data to detect data. There are two type of recognition method which is hard Sensing and soft sensing. In hard sensing, sensor can detect emotion like brain signal, heart rate, eye gaze etc, while in soft sensing sensor can detect emotion from software like, email, text messaging, desktop, social

interaction etc. In this report we will pick out text as text emotions because the communication is between human and machine. In a survey report 68% [] people use text messaging to family or friends while only 49% people use face to face communication. From face to face talking you will find person impression directly while in text messaging this is not such easy to find the person impressions, so the computer need such kind of algorithm which can detect emotions. In this survey report we will use Emotion Recognition rules through sentence level. Here **split a sentence on the basis of Noun, pronoun, verb, adverb and some phrases and then find the emotions**. Sentence like "Hooray we won the match." Now there is relationship between noun "Match" and verb "won".

The motivation behind this survey report is the rapid growth of World Wide Web give online communication and provides a gate way for public to post their opinion, emotion. We need computational algorithms to analyze this large amount of data and draw some useful conclusions. Recognizing text is also useful for psychologist to analyze the person views.

### **Applications:**

There are different Applications from mood detection through text. Some of the applications are as under.

### **Sentiment Analysis:**

Sentiment analysis also called opinion mining, it focus on retrieval of information from the text. Now computer should be able to detect useful information from text. It has advantages like to collect useful information from blogs, website and through these information decision taken that either the customer is happy or not. Sentiment analysis is useful for online shopping where to find useful reviews of customer and through these reviews to decide whether customer is happy or not.

### **Text to speech:**

Document consist of many sentences and the reader read through different way like some time increase the pitch, volume and sometime different languages people pronounce through different ways. The text to speech generation is an new era application which we enter to the modern world. And through this application you convert your text to exact speech to pronounce that word.

### **Human Computer Interaction:**

The Mood detection through text has high advantage to modern world. Through this way you can increase the knowledge of robot and find the emotion of robots, it also works in automatic answering, and in dialogue system. Sometime when sending an email the email server replies you automatically.

### **Individual Consumers:**

If a person want to purchase a product from a company and the person just go the website and instead of reading all the customer reviews and wasting time of reading the customer just to understand the product usage and usefulness through only one emotion.

### **Recognition Unit:**

This survey will focus on emotion recognition in a sentence. Because it is difficult to find

emotions in sentence rather than the case where there is already an emotional word. It is easy to find emotions in a text documents, paragraph etc there may contain many emotions. If the recognition unit is small like keyword it is easy to find emotion but in sentence it may have one intension at a time. We also found such kind of emotions detection system in robots, chatting system etc.

### **Related Work Survey:**

There are a lot of research works in the field of Emotion detection through text; Different people use different sentence to express their feelings like "Oh my God you make this program" here the sentence

show some surprise emotions. Emotion can be detected through formula let say, E set of possible events, A set of authors, T representation of emotion in text and r be the function to reflect emotion e from author a it is formulated as i.e.  $r: A * T \rightarrow E$ . May be it could be the best but the problem is new emotion is added day by day and there is no specific standard for emotion like recently in 2015 it say there are only 4 emotions. Here we classify emotions into two categories coarse grained and fined grained. **Hancock in (2007)** used linguistic inquiry and word count (LIWC) analysis and classifies emotions into positive and negative. Negative emotion expressed in more effective words and positive expression expressed through exclamation mark. But the disadvantage of this method is, that is limited for only happy and sad emotions. Different persons show different states of emotion and classify emotion. Here different persons show different models to classify here the words.

Drummond	Ekman	Izard	Plutchik
Anger	Anger	Anger	Anger
Caring	Disgust	Contempt	Anticipation
Depression	Fear	Disgust	Disgust
Fear	Happiness	Fear	Fear
Happiness	Sadness	Guilt	Joy
Hurt	Surprise	Interest	Sadness
Inadequateness		Joy	Surprise
Loneliness		Sadness	Trust
Remorse		Shame	
		Surprise	

Fig 1: Different proposed model

Now to classify emotion in fined grained we use six ekman's emotions and that can be done with through three different methods Keyword based, knowledge based, Hybrid Based.

## Test Based Emotion Recognition Method:

In the previous work they show only 3 recognition method while here in this survey we will show 4 recognition methods.

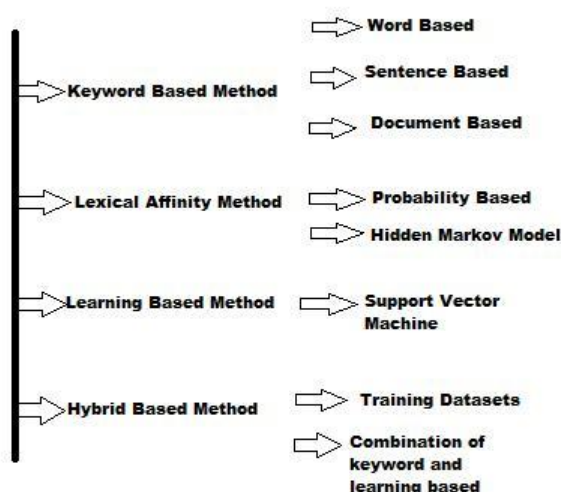


Fig2: Emotion recognition methods.

## Keyword Based Approach:

This method is traditional method and very easy to find emotional keyword, it depends on input sentence. In the early times Osgood in 2007 use three dimensional method to identify emotional

scale which are **potency, evaluation and activity**. **Evaluation** shows that **how much a word is close to**

happy or sad, it show the sentence negativity and positivity, Potency shows strong and weak intensity of word and activity shows the passive or active activity of the current sentence or word.

To identify specific word in the text, apply some matching patterns to check the whole sentence for finding a specific keyword. The words are “disgust, sadness, happy, anger, fear and surprise”.

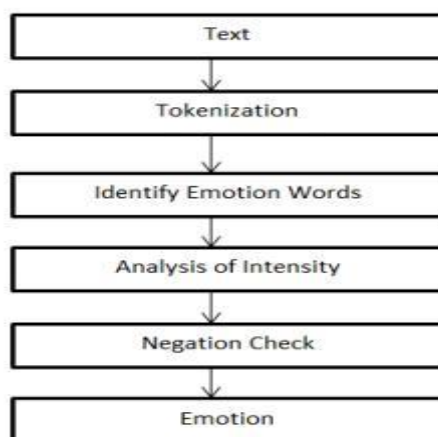


Fig3: Splitting Text to emotion

Now to clear this input must be text document while output must be an emotion. Take the input as document and applying tokenization rule to convert into token emotional keyword. Now perform some analysis on emotional keyword to find the intensity of word and after this check the whole sentence for negation if found or not then the emotion class will be found here.

## Machine Learning Based Method

Working through machine learning methodology this face some problems, split the sentence into two categories positive and negative. Look to the previous research like Strapparava (2008) used a system for text when there is no effective word exists but this system gives us low accuracy if there is no specific or effective word in the sentence. This system was neither context nor semantic sensitive. Dung in [2013] proposed idea that events are related to human emotions. This means that human mind events changes from one state to another according to some emotional event, But this was also non context sensitive. This idea was implement in HMM (Hidden markov Model).

### Hidden Morkov Model:

HMM work on probability and work only on two emotions happy or Sad, and how much the probability that person is happy or not. HMM work in this way like future is independent of the past given present. This mean that you want to predict for future from the past and there is some event which already exist in present, but HMM solve the problem and it says that may be sentences consist of sub idea and it could be event in the system.

### Support Vector Machine:

This algorithm classify sentence into vector categories. This algorithm makes a matrix, where row is word and column is a sentence, now the row represents term and relationship while column vector represent document and their relationship. Now VSM use frequency schemas which are tf-idf, tf stand for term frequency and find count the number of times in the document a word appear, and idf stand for inverse document frequency and show the word is rare or common in the document.

$$tf = nt, d / kd$$

Number of time a word t appears in a document d and kd is the total number of times in the document.

$$idf = \log_{10} * |D| / |Dt|$$

Where D is the total number of document in the corpus and Dt is the number of documents where t



appears.

The binary hyperlane is between the features of data to between positive and negative. To select the perfect hyperlane we must differentiate through margins between the data. Through this way we can find maximal hyperlane between the different data.

### **Decision Tree:**

Decision tree is a binary tree which is represented by nodes, tree work in recursive algorithm. Each node has two child nodes and the decision is made before the leaf node. Let say we identify a certain data D, take the training data and divide until the data d found in the leaf nodes. All the information gain from the leaf nodes, finds the probability at each node, and move to that way where there is higher probability of finding data.

### **Naive Bayes Algorithm:**

Byes Algorithm work based on probability to move from Unconditional probability to conditional probability, it says based on condition there might be an event [3].

$$P(G|U) = (P(U|G)P(G))/P(U)$$

And

$$P(G|U) = (P(U|G)P(G)) / (P(U|G) + P(U|G'))$$

$P(G|U)$  This means that probability of G occur when U already occur.  $P(U|G)P(G)$  called joint probability This means that  $P(G)$  occur when the  $P(U|G)$  when G occur.

### **K-Nearest Neighbor (KNN):**

KNN is a simple algorithm and it could be solution for any classification problem. KNN basically work on similarity basis or distance basis to find the nearest possible points. The

points which are going to be compare is K points. This algorithm picks all the nearest points based on nearest distance but this algorithm is very slow because  $O(N)$

### **Lexicon Based Method:**

The keyword based method is a straight forward method where to detect specific keyword, But to find a keyword on the basis of probability the sense to take positive or negative. Some lexical meaning of word is different as use of the sentence. Now for each emotion type, count the number of emotion in each sentence. Here Support Vector machine play an important role for classification. The term frequency and inverse document frequency show number of time a word appear in a count. For example the word accident show negative emotion but it depends on a sentence like how to use this word,

*"I met my brother by accident"* Now there is probability that the sentence show negativity because of the word accident.

### **Hybrid Based Method:**

Hybrid Approach is the combination of Keyword based approach and training datasets. Whenever the system receives the input and check the text whether it has keyword emotion or not. If the input text has one or more emotional keyword then we use keyword approach. Now determine which keyword is used here we use Emotional keyword dictionary, where we have emotional word meaning if word found in the dictionary then we have specific output of the system. If there is no emotional keyword in the document then we will use emotion classifier method.

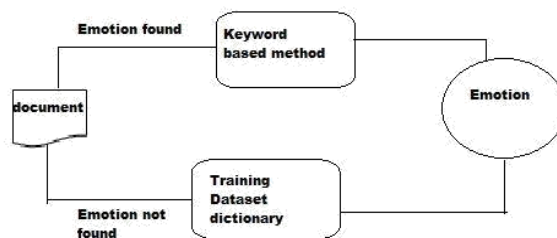


Fig 4: Methodology for Hybrid Model

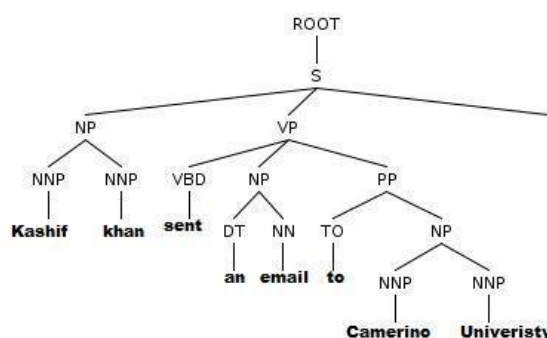
## Mathematical or Technical Solution:

In this survey we will apply some technical techniques on English sentence to reduce restrictions as we have study in different research papers. Now  $T$  is test and  $s$  is sentence so  $s \in T$  let  $E=[empos, emneg]$ ,  $empos$  this mean positive emotion and  $emneg$  negative emotion. Now the goal is how to find mapping function  $f:s \rightarrow E$  now to get  $(s, emn)$  where  $n$  is pos or neg emotion, for every sentence or in a sentence we use  $f=\{f_1, f_2, f_3 \dots f_n\}$ . There is specific and unique concept in each and every sentence and through these concepts we extract emotions. Compositionality principle [4] states if the Lexical (meaningful) part is taken out the remaining is composition rules. For example: “*I solve this problem*”. This sentence shows some happiness, and the emotion could be joy. Now the main words(meaningful) are “solve” and “problem” now convert it to pseudo code like “I S the P” (S for solve and p for problem) So the syntax and semantic relation is between the solve and problem.

Second principle is sentence construction principle; in every sentence, analysis could be done through semantic and syntax level. Through semantic level the sentence must be understandable and meaningful. Through syntax analysis can be done through part of speech. “I found the same problem but it has different solutions”. In this sentence PRP, VB, the, ADJ, but, it, has, ADJ, NN, here PRP is pronoun, VB is verb, ADJ is adjective, NN is noun. Now to

look at the semantic level there is relation between the PRP I and VB found and on the other side VB found with NN problem and NN solution. Now here is the strong relation between syntax and semantic, we cannot ignore importance and their relationship. Now to make a rule, effected words can be nodes while other can be sub nodes, An Idea take place to make the sentence in the form of tree so representation of can be done through Emotion Recognition Rules(ERR).

To Constructed ERR we will use 4 types of rules here. We have adverbs, noun clauses, verb-noun clauses, and adjectives. Now here we use some dependency trees which are parser tree [6]. I am writing a sentence and then convert it into tree base like “Kashif khan sent an email to Camerino University” so here we have “N,N,V,T,N,N,N”  
So here is the relationship.



Here NP means Noun phrase and VP mean verb phrase. Now the parser for this tree is :



Parser

(Root

(S

(NP (NNP kashif) (NNP khan))

(VP (VBD sent)

(NP (DT an) (NN email))

(PP (TO to)

(NP (NNP camerino) (NNP university)))))).

Through this base we can find algorithm for sentence which have POS.

**Data** take the sentence

Start with root node

**For** every sentence **do**

Extract into NP and VP

**For** NP **do**

Extract into POS

Find noun

**END**

**For** VP **do**

Extract into POS

Find events

Find verb

**END**

Repeat until all the phrases split

**END**

**Result:** Find number of nouns, Verbs, events.

Now apply Stanford Parser and we will apply rules to pick the emotional part from the sentence.

**Rule 1:**

They rule say ignore the complete sentence before word "BUT" like we have a sentence. "We try our best to complete our work but it was difficult". Now remove the sentence "we try our best to complete



our work” now the remaining part is “it was difficult” now from here you can identify emotions.

**Rule 2:**

Ignore the complete sentence or phrase after the word “as”. The word “as” work as conjunction for two sentence or words phrase. In a sentence like “He is good as his father”. So remove the sentence after as like “his father”. So the remaining is “He is good” from this we can identify emotions from a sentence.

**Rule 3:**

Some time remove the Verb from the sentence to pick the emotional word. Like “we had fun” now remove the word “Had” and the relationship is in between the word we and fun.

**Rule 4:**

Remove WP (Who, what, which etc) pronoun if it has no relation with the previous sentence.

“What are you doing here it is not a good place”.

Now remove the sentence which has WP pronoun (What) what are you doing here. So the remaining part is “it is not a good place’ the last sentence shows some emotions here.

**Restrictions:**

There are a lot words which has different meaning at different stages. But there are some major challenges which would be solvable in the future.

**Ambiguity of word:**

Most of the same words has different meaning at different stages but this a major problem that how to handle this. Like right, “*The professor is right. /Go downstairs and then turn right*”, so same word but different meaning in different sentences.

**Sentence with no emotional keyword:**

Our approaches would be wrong if there is no emotional keyword or no emotional feelings then how to handle such kind of sentences. Like sentence “Hooray I passed my Database exam” if we remove hooray which is emotional word then how to handle such kind of sentences.

**Lack of Linguistic Information:**

Lack of linguistic information is also a problem that you cannot judge a sentence through syntax and semantic information. In English language sentence the usage of syntax and semantic have different meaning at different sentence like “he laughed at me” and “I laughed at him” Now the meaning is completely different at these two different sentences.

**Critical Evaluation:**

Emotion from text is critical part of research areas. We check different approaches on sentence level and pick the best approach to solve the problem. So here to use the key based approach, this approach will pick only a specific word which has emotion in the sentence but it create a problem that when a word in a sentence have different as compare to original meaning. Lexical approach work on probability approach and some time it picks the world which has high probability instead of emotional world, due to this drawback we do not use this approach. Hybrid approach work on dataset or dictionary and is the combination of lexicon and keyword based method, the accuracy is 7 times more than that of other true but the disadvantage is when we have a emotional word which are not present in the dictionary. Machine learning approach is in the developing phase, many researchers are working on this approach to make a perfect algorithm.

As comparison of the previous researches and this current research is that we include the some

algorithm like SVM, Byes algorithm, Hidden Morkov model to the previous approaches. Here we also take the Stadford Parser and make a new algorithm on the basis of parser tree. We also make some rules on the basis of part of speech and split the sentences in token form to make the correct system to identify the emotion in a sentence or document level.

### Conclusion:

Emotion detection is an important part in the field of Human Computer Interaction. Due to high usage of social networks or online shopping many researcher want to work in this field. To classify emotion we need some quick algorithm to detect emotion instead of reading all the websites or blogs.

### References:

- [1]: P. Ekman, (1999) Basic emotions. In T. Dalgleish and T. Power (Eds.) *The handbook of cognition and emotion*.
- [2]: A Survey of Spelling Error Detection and Correction Techniques Ritika Mishra<sup>1</sup> , Navjot Kaur<sup>2</sup>
- [3]: Byes law [https://en.wikipedia.org/wiki/Bayes%27\\_theorem](https://en.wikipedia.org/wiki/Bayes%27_theorem) ... (Online)
- [4]: [https://en.wikipedia.org/wiki/Principle\\_of\\_compositionality](https://en.wikipedia.org/wiki/Principle_of_compositionality) .. [Online]
- [5]: <http://www.link.cs.cmu.edu/link/submit-sentence-4.html>..[Online]
- [6]: <http://nlp.stanford.edu/software/stanford-dependencies.shtml#Methods> ... [online]
- [7] Dependency-Based Semantic Parsing for Concept-Level Text Analysis ( Soujanya Poria<sup>1,4,6</sup> , Basant Agarwal<sup>2</sup>)
- [8] Study and Review of Various Emotion Classifications from text techniques by (Sandip S. Patil and Chaitali G. Patil)
- [10]: "Text Based Emotion Recognition: A Survey" **Chetan R. Chopade**
- [11]: "A Review on the Emotion Detection from Text using Machine Learning Techniques" Ubeeka Jain<sup>†\*</sup> and Amandeep Sandhu<sup>†</sup>
- [12]: "Emotion Detection from text: A Survey" Lea Canales, Patricio Mart'inez-Barco
- [13]: "TexEmo: Conveying Emotion from Text-The Study" Mukesh C. Jain
- [14]: "Analysis and Identifying Variation in Human Emotion Through Data Mining" Jasakaran Kaur
- [15]: "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu"
- [16]: "TextTone: Expressing Emotion Through Text Ankur Kalra and Karrie Karahalios".
- [17]: "Rezabek, L. L., & Cochenour, J. J. (1998) Visual cues in computer-mediated communication: Supplementing text with emoticons. Journal of Visual Literacy, 201-215."