

# Survey on Emotion Detection from Text using Automatically Generated Rules

Anuradha V. Yenikar, Simrit Raina , Ashlesha Kothawale , Shruti Mate  
Zeal College of Engineering and Research

**Abstract:** Feelings have a noteworthy job in everyday correspondence between two people. Feeling identification can be coordinated and formed into the content to discourse applications. Most specialists have endeavored to identify the client's feeling in some ways, for example, outward appearance, voice, and content. A fruitful intelligent framework between individual and PC can be accomplished when the framework can perceive & decipher the outflow of a person feeling precise. The user emotions are recognizing in two way i.e. hard sensing and soft sensing. We can analyze the user by their gestures, emotions, and thoughts. An emotion can be described as a state of mind to express our feelings which affection, rage, excitement etc. Programmed feeling recognition from content has pulled in developing consideration because of its possibly helpful application. Perceiving feelings through content is one of the greatest difficulties for people and machines. The feeling location framework can be connected to numerous areas also, for example, in business (to perceive the clients' impression dependent on their announcement on the item offered), in training (used to identify the understudies' excitement for a web-based learning framework), in PC gaming, in emotional wellness, in national barrier, and others. Distinguishing human feeling through content is considered as the most basic one.

Keywords: Emotion types; Textual Emotions; ERR; Emotion analysis.

## I. INTRODUCTION

Feelings or emotions are an integral part of our day to day life. Human mind can have variety of emotions which according to researchers have been categorised into 27 types but for basic it has categorised into sad, anger, disgust, fear, joy, happy. How someone is feeling can describe his/her activities and behaviour towards others. Research have suggested that in order to know emotions better it can be divided into three elements the subjective experience, the physiological response, and the behavioral response [1].

### The Subjective Experience

Even if the emotions are classified into some of the basic type but they all can be felt in many ways depending upon the situation and position. For e.g. a joy of getting passed in an examination is way different from that of joy of becoming parents [1].

### The Physiological Response

Research have shown that there are many emotions that are controlled by our nervous system known as sympathetic nervous system(SNS) [1]. Emotions like butterflies in stomach or heart sinking are some of the examples of SNS. SNS can also result in sweaty palms or heart beating getting faster. These types of emotions come under the physiological response.

### The Behavioral Response

Humans always have a tendency to know the emotions of the persons he is surrounded with or interpreting what the other persons state of mind is. Emotional Intelligence can be described as the ability to accurately figure out the emotion of a person. This expression play an important role in knowing overall body language of a person [1].

Detecting emotions has been always a topic of discussion among researchers and a sufficient research has been done in detecting human emotions through facial expression and through speech. Detecting human emotions through text is still a new field and as compared to facial and speech detection. Detection of emotions through text can help in commercial world as well as in education field. Previous works has been done in regarding with text classification which can help to know sentences and paragraphs better and extract the emotion in it. Retrieval of information through text comes under the concept of text mining. The mining of text includes three processes: (I) The input text should be properly structured, (II) After applying preprocessing steps like removing stopwords, parsing patterns are derived (III) The retrieval of required information is the last step [2].

Classification of text is considered to be a field that comes under text mining. With the rapid growth of social media and its use among people, recognizing emotions have become very important for the machines in order to act more smartly and intelligently according to mood of the user. Researchers have used many classification algorithms such as KNN, SVM, Naive Bayes, Decision Tress, Random Forest in order classify texts and extract the emotion in it.

## II. LITERATURE SURVEY

### Keyword-Based Detection Method

Past work done on understanding feeling articulation in content was finished by **Osgood et al.** [3, 5]. Multidimensional scaling is utilized by them for picturing the full of feeling words so as to compute similitude evaluations between them. Osgood utilized three measurements, for example "assessment", "movement" and "strength", where assessment can gauge how much a word alludes to a wonderful or an undesirable occasion, enthusiastic force of a word (solid or frail) is measured by power, and action alludes to the dynamic or detached nature of a word.

**Torao Yanaru.**, built up a phonetic asset for a lexical portrayal of full of feeling learning named WordNet – Affect [13]. In this strategy sentences and sections are examined via scanning for enthusiastic catchphrases like that of delight, dread, outrage, trouble, disturb. This technique utilizes lexicons like WordNet to check the similitude between the enthusiastic watchword found in the sentence to that of its word reference which is a mix of numerous comparative passionate words and on the off chance that any likeness is found between the two words, at that point, it creates the sort of feeling.

**Santanu Mandal**[7] the Keyword based technique is a straight forward strategy where to identify explicit catchphrase, but to discover a Keyword based on the likelihood sense to take positive or negative. Some lexical importance of word is diverse as utilization of the sentence. Presently for every feeling sort, include the quantity of feeling each sentence. Here Support Vector machine assume an imperative job for order. The term recurrence and backwards record recurrence show number of times a word show up in a check. For instance, the word mishap shows negative feeling however it relies upon a sentence like how to utilize this word, "I met my sibling unintentionally" Now there is a likelihood that the sentence show negativity due to the word unintentionally".

### Machine Learning Based Method

**Strapparava** [12] built up structure so as to use differing assortments of Latent Semantic Analysis in order to recognize sentiments within printed data while no of the emotional words exist. Anyway because of it's not being setting touchy and lacking of semantic examination of the sentence, lower exactness is accomplished by the framework.

**Ghazi et al.** [8] endeavored different leveled request to mastermind the six Ekman feelings(emotions). They used different components of a levels of leadership while portraying sentiments by first gathering paying little respect to whether a sentence holds an inclination or not, by then assembling the feelings(emotions) are named positive or negative in end engineering the feelings on a well grained estimation. By using assorted features at every period over classifier they accomplished a predominant precision (+7%) above dimension request where level portrayal is orchestrating feelings on a well grained measurement explicitly. The major setback to this method was its lack of sensitiveness.

### Hidden Markov Model

**Dung et al.** [2] proposed thought that occasions are identified with human feelings. This implies human personality occasions changes starting with one state then onto the next as per some enthusiastic occasion, But this was likewise non setting touchy. This contemplation was realized in the HMM in whichevery phrase involves diverse subordinate-musings as well asevery contemplation is seen as an occurrence which rolls out an improvement in a particular state. The system chooses the most conceivable sentiment of the substance, by following the game plan of events in the sentence. Well work on likelihood and work just on two feelings cheerful or Sad, and how much the likelihood that an individual is upbeat or not. Gee work thusly like future is free of the past given the present. This implies you need to anticipate for future from the past and there is some occasion which as of now exist in present, however, HMM tackle the issue and it says that might be sentences comprise of sub thought and it could be an occasion in the framework.

## Hybrid detection

**Cheng-Yu [10]** displayed an emit point artistic inclination, identifying with construction of shared histogram among pair of components in which every portion of histogram addresses as fundamental movement (activity word) existed between the two substances. A F-score of 75% was practiced when attempted on four sentiments. In any case, technique fails to prove significance in sentence as well as is exceedingly reliant on arrangement of data and its structure, for instance, sentences having semantics within planning data as well as repeat of affections meant in support of particular matter. Ekman's four emotions were utilized in this gathering out of six.

**Wu [11]** put forward a methodology for drawing out emotions from sentence point depending distinguishing over recently characterized Symantec names as well as characteristics in sentences, at that point characterization of emotions dependent on mental examples of human feelings called as feeling age rule (EGR). However, their methodology was confined to one feeling (cheerful) since the technique demonstrated a great deal of vagueness when one EGR can produce more than one feeling.

## Lexicon Based Method

**Neviarouskaya [9]** created Emhart, a lexical decide framework so as to perceive feelings commencing content and envisions inside practical situation. The framework begins through searching meant for enthusiastic shortened forms as well as emojis. Unless discovered, it forms verdict on diverse dimensions (wordpoint, express dimension as well as sentence point) in order to create passionate vectors, where every component of vector speaks to passionate group power. On wordpoint, every word in sentence is mapped on its corresponded enthusiastic vector, in which enthusiastic vectors of some words are physically assembled as a dataset. On expression as well as sentence point, there is joining of enthusiastic vectors which are collected either by performing amplification or summation in between the vectors. They accomplished a normal exactness of 75% when tried on a physically explained dataset. Anyway this strategy displays scarcely any downsides. To begin with, the framework does not deal with the case at the point when invalidation exists in the sentence. Second, it depends on a full of feeling database where feeling classifications and forces were appointed physically to each word in the database, which makes their methodology difficult to stretch out to arrange more feelings.

## K-Nearest Neighbor (KNN)

**Aiman Moldagulova [4]** proposed KNN, KNN is a basic calculation and it could be answer for any characterization issue. KNN fundamentally deal with comparability premise or separation premise to discover the closest conceivable focuses. The focuses which will be compared are K focuses. This calculation picks all the closest focuses dependent on closest separation however this calculation is exceptionally moderate since  $O(N)$ .

## Decision Tree

**Y. Sakakibara [6]** choice tree (Decision Tree) is a twofold tree which is spoken to by hubs, tree work in a recursive calculation. Every hub has two type hubs and the choice is made before the leaf hub. Let say we distinguish a specific information D, take the preparation information and separation until the information d found in the leaf hubs. All the data gain from the leaf hubs finds the likelihood at every hub, and move to that route where there is a higher likelihood of discovering information.

### III. PROPOSED SYSTEMS

#### 1. Load Textual Data:

In this phase, the user can upload the textual data for emotion detection. They can load any sentence, paragraph etc. as an input.

#### 2. Preprocessing:

In this phase, the user can apply preprocessing on the input data. Preprocessing includes,

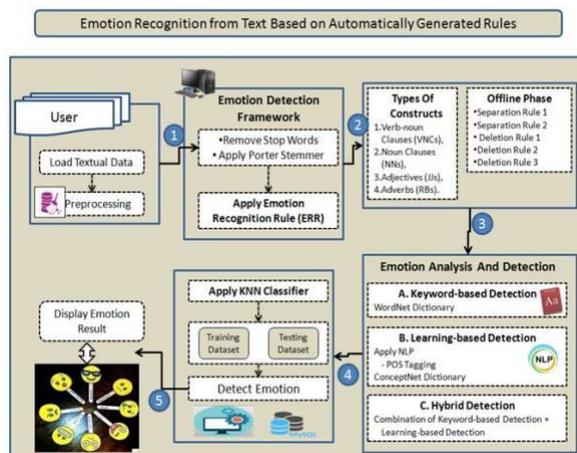
- Remove Special Symbols (@,#,&,etc.)
- Remove Stop words
- Apply Porter Steamer

#### 3. Apply Emotion Recognition Rule (ERR):

The ERR are comprised of following

- 1) Verb-Noun Clauses (VNCs),
- 2) Noun Clauses (NNs),

Figure 1. System Architecture



- 3) Adjectives (JJs), 4) Adverbs (RBs).

#### Offline Phase:

In this stage, the framework can apply the accompanying guideline on sentence in order to generate some rules which will later work as automatically generated rule,

#### 1. Separation Rule 1:

In the event that the sentence contains "but" word at that point disregard the sentence before "but".

#### 2. Separation Rule 2:

In the event that the sentence contains "as" word at that point overlook the sentence after "as".

#### 3. Deletion Rule 1:

Evacuate an actionword on the off chance that has not any article as well as is associated with a WRB or WP pronoun fore.g. "Where you are going is a sickening spot?" The part "where you are going" is expelled and the latter half of the sentence i.e. "sickening spot" will act as ERR.

**4. Deletion Rule 2:**

Evacuate an action word hub on the off chance that it is either non-enthusiastic foreg. "We had a ton of fun" is limited to hold just two hubs "we" along with "fun".

**5. Deletion Rule 3:**

Evacuate pronoun hubs on the off chance that they are not associated with different hubs. For example, "It was somewhat convoluted yet we had a pleasant time", the main hub has "pleasant" in it and is going to work as ERR[12][13].

**4. Emotion Analysis and Detection:**

The emotions can be examined by utilizing three strategies,

**A. Keyword-based Detection:**

In this strategy, the feelings are distinguished by utilizing passionate(emotional) Keywords. They likewise utilized the WordNet lexicon for finding the comparable importance words. WordNet Dictionaries- WordNet is a mix of lexicon and thesaurus, what it does is that it puts words of English into a group of corresponding words which are to be known as synsets. Synset work is simple as it provides simple meaning to words by using various models and then finds a relation between these corresponding set of words.

**B. Machine Learning based Method:**

In this technique, the feeling is identified by utilizing the preparation dataset and afterward grouping methodology. They can apply NLP preprocessing on the sentence or word.

**Apply NLP-**

Natural language handling (NLP), a part of computerized reasoning(artificial intelligence) concerned about robotized understanding and age of human language. It is a cosmology helped method for programming as far as regular language sentences.

**POS Tagging -**

Part-Of-Speech (POS) recognizes whether the word token is a thing(noun), an action(verb) word, and modifier(adjective). The word is doled out as per its syntactic functions. In English, the principal parts of discourse are a thing(noun), pronoun, modifier(adjective), determiner, an action word(verb), relational word, intensifier, combination, and contribution.

**Calculate TF-IDF** –Compute significance amideverywordplus every feeling set.

**C. Hybrid Detection:**

It is a mix of both above strategies, for example, Keyword based location and learning-based recognition.

**5. Apply KNN Classifier:**

The KNN classifier is used for to detect emotions. Firstly, they computing the similarity between two ERR data. For computing the similarity of each ERR construct they use the Word Net, Concept Net dictionary. Also they can compare VNCs, nouns, adjectives/adverbs list of both ERRs.

They can use the following formula for calculating resemblance,

$$\text{resemblance} = \text{Res}(\text{Verbnoun Clauses}) + \text{Res}(\text{NounClauses}) + \text{Res}(\text{AdjectiveClauses}) + \text{Res}(\text{AdverbClauses})$$

Also, the system can apply the backup classification methods, like PMI, PMI-IR.

**PMI,**

$PM(F; W) = \log(\frac{\#part\ with\ feeling(emotion)F\ having\ word\ W}{\#whole\ sentences\ having\ W})$

**PMI\_IR:**

$PM\_IR(F; s) = \log(\frac{\#Retrieved\ hits\ for\ query\ (s + F)}{\#Retrieve\ hit\ for\ the\ query\ (s)})$ .

## 6. Detect Emotion:

In this phase, the final emotion detection result is display to the user. The system can notice six types of emotions like,

- Happiness
- Sadness
- Surprise
- Disgust
- Anger • Fear

## IV. Algorithm

### 1. K-nearest(closest) neighbors (KNN) Classification Algorithm

Ventures for processing K-closest neighbors KNN calculation:

1. Determine parameter K = number of closest neighbors.
2. Calculate the distance(separation) between the inquiry case(query-instance) and all preparation tests.
3. Sort the separation(distance) and decide the closest neighbor dependent on the k-th least separation.
4. Gather the classification of the closest(nearest) neighbors.
5. Use basic dominant part of the class of closest neighbors as the forecast estimation of the inquiry occurrence.

### 2. Porter Stemmer Algorithm:

In phonetic(linguistic) morphology and data recovery, stemming is the way toward diminishing arched (or in some cases determined) words to their pledge(word) stem, base or root structure—for the most part a composed word structure. Following are the steps of this algorithm are:-

- i. Gets rid of plurals and -ed or -ing suffixes
- ii. Turns terminal y to i when there is another vowel in the stem
- iii. Maps double suffixes to single ones: ization, -ational, etc. iv. Deals with suffixes, -full, -ness etc.
- v. Takes off -ant, -ence, etc.
- vi. Removes a final -e

## V. CONCLUSION

It has seen that numerous past works and models have come in forward so as to perceive feeling through content or setting. Be that as it may, perceiving feeling through content is as yet a field in which new strategies are been risen and contemplated. In this paper, we have experienced a portion of the procedures and strategies that are there altogether to perceive feeling through content. While examining we have discovered that the downsides of the past models, for example, uncertainty in words, the absence of perceiving sentence without catchphrase and absence of phonetic data. We have proposed a model which is going to help in perceiving feelings from writings utilizing automatically generated rules. This model uses Keyword based, learning based and hybrid methods so as to perceive feeling behind any specific situation. The upside of the proposed framework is that it will conquer the downside of vagueness in words and perceiving sentences without a catchphrase.

**REFERENCES**

- [1] Kendra Cherry , "Emotions and Types of Emotional Responses The 3 Key Elements That Make Up Emotion"
- [2] Hema Krishnan, M. SudheepElayidom, T. Santhanakrishnan "Emotion Detection of Tweets using Naïve Bayes Classifier", 2017.
- [3] Jeffrey Cohn, and Gary Katz. "Bimodal expression of emotion by face and voice."IEEE 2017
- [4]Aiman Moldagulova,"Using KNN algorithm for classification of textual documents" IEEE 2017
- [5] Dung T. Ho, and Tru H. Cao. "A high-order hidden Markov model for emotion detection from textual data.", IEEE 2017.
- [6] Y. Sakakibara , "Text classification and keyword extraction by learning decision trees" IEEE 2016
- [7] Santanu Mandal , "A Lexicon-based text classification model to analyze and predict sentiments from online reviews" IEEE 2016
- [8] Diman Ghazi "Hierarchical versus flat classification of emotions in text." NAACL, 2010.
- [9] Alena Neviarouskaya, Helmut Prendinger"EmoHeart: conveying emotions in second life based on affect sensing from text.",2010
- [10]Cheng Yu Lu, et al. "Automatic event-level textual emotion sensing using mutual action histogram between entities." IEEE 2010
- [11]Chung-Hsien Wu, Ze-Jing Chuang "Emotion recognition from text using semantic labels and separable mixture models." 2006
- [12]Anuradha V. Yenikar, Sachin S. Pande, "Rule Based System for Product Lifecycle Management", IJCA, Volume 59– No.19, pp. 33-37, 2012
- [13]Anuradha V. Yenikar, Sachin S. Pande, "PLMRBS: A Forward Chaining Inference Engine Optimized for Product Lifecycle Management Business Rules", IJARCSSE Volume 3(4), pp. 110, 2013
- [14]S. M. Kolekar, P. N. Mahalle," Malware Prevention and Detection System using Smart Phone"International Journal of Computer Applications (0975 – 8887) Volume 107 – No. 21, December 2014
- [15]Sachin M. Kolekar, Rahul P. More, Smita S. Bachal, Anuradha V. Yenikar, "Review paper on untwist Blockchain: A Data Handling Process of Blockchain Systems", ICICET, Volume No. 18 Issue: 978-1-5386-5510-8, pp. 1-4, 2018.
- [16]Carlo Strapparava "WordNet Affect: an Affective Extension of WordNet." *LREC*. 2004.
- [17]ToraoYanaru, "An emotion processing system based on fuzzy inference and subjective observations." IEEE, 2002
- [18]Zahra Khalili "Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG." *IJCNN*, 2000.