

PAPER • OPEN ACCESS

## Emotion detection from text

To cite this article: V V Ramalingam *et al* 2018 *J. Phys.: Conf. Ser.* **1000** 012027

View the [article online](#) for updates and enhancements.

### Related content

- [Facial recognition in education system](#)  
L B Krithika, K Venkatesh, S Rathore et al.
- [Emotion models for textual emotion classification](#)  
O. Bruna, H. Avetisyan and J. Holub
- [Physics as a humanity](#)  
R C Chivers



**IOP | ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

# Emotion detection from text

V V Ramalingam<sup>1</sup>, A Pandian<sup>2</sup>, Abhijeet Jaiswal<sup>3</sup> and Nikhar Bhatia<sup>4</sup>

<sup>1,2,3,4</sup>Department of Computer Science and Engineering, S.R.M. Institute of Science and Technology, Chennai, India - 603203

*ramabi1976@gmail*

**Abstract.** This paper presents a novel method based on concept of Machine Learning for Emotion Detection using various algorithms of Support Vector Machine and major emotions described are linked to the Word-Net for enhanced accuracy. The approach proposed plays a promising role to augment the Artificial Intelligence in the near future and could be vital in optimization of Human-Machine Interface.

## 1. Introduction

Emotion Detection will play a promising role in the field of Artificial Intelligence, especially in the case of Human-Machine Interface development. For Emotion Detection from an artificial intelligence different parameter should be taken into consideration. Various types of techniques are used to detect emotions from a human being like facial expressions, body movements, blood pressure, heart beat and textual information. This paper focuses on the emotion detection from textual information. Nowadays within the Internet there's an immense amount of textual data. It's fascinating to extract emotion from various goals like those of business. As an example, in luxury merchandise, the emotion aspect as brand, individuality and prestige for purchasing confirmations, are lot necessary than other aspects like technical, functional or price. In such conditions buyer is happy to shop for a product even with high costs. Emotion selling aims to simulate emotions in clients for tying them to brand and then increase the selling of service/product. While nice strides were created as in emotion recognition exploitation multimodal sources, such as: face, voice or gestures, there's not yet a strong enough text - based feeling recognition solution, capable of detecting emotions from text, with high accuracy, in spite of the text size, and taking into consideration context or one's type of expression. There are four basic methods to detect emotions from text: 1) Keyword-based detection, 2) learning-based detection, 3) lexical affinity method, hybrid detection. Each and every method contains some strong and weak points while detecting emotions from text. Hybrid Method is the most likely method to get a high accuracy result, as it includes the strength of combined strength of two or more methods. In that also main difficulty is to find the most effective combination.

In all the methods these challenges generate problem related to emotion detection:

- Collection of Data: what data should be used for feature extraction? And how to cope up with the continuous changes or evolution of textual expressions used in everyday exchanges?
- Features Choices: which type of emotion indicators can be present in a speech? How contextual data can be extracted? How to combine those features to get a good result?
- Labeling of Emotions: what emotions are going to be assigned in a piece of text? Especially



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

in the case of multiple word combination. And what categories of emotions to be used for the training dataset?

- Machine learning classifier: What is the best classifier to use for various textual data? More than one classifier should be used?

In this paper, we have tried to solve such problem and introduce a method to get high accuracy result while detecting emotion from text. We have combined both the keyword-based and learning-based method to get the result.

## 2. Emotion Detection Methods

Emotion detection approaches use or modify concept and general algorithm created for subjectivity and sentimental analysis. There are many approaches that are being used and explored. However, many of the approaches have few similarities in them. Some of the methods available are presented here.

### 2.1 Keyword-based Methods

Keywords based approaches use synonyms and antonyms are WordNet to determine word sentiments based on a set of seed opinion words. In a bootstrapping approach is proposed, which uses a small set of given seed opinion words to find their synonyms and antonyms in WordNet to predict the semantic orientation of adjective. In WordNet the adjectives are in bipolar cluster form of organization and have synonyms have same orientation. As all the adjectives are linked and it form a pattern and leads to the emotion which the word depicts.

### 2.2 Vector Space Model

Categorical classification is used in the approach of Vector Space Model (VSM). Matrix of co-occurrence frequency vectors are used to representing the dataset dimensionally. Words are represented by rows and the columns can represent sentence, paragraph or documents. Therefore, the column and the row depict a relationship. VSM weighs these frequencies using the tf-idf weighting schema.

The tf-idf score is the weight of each word in terms of its importance within the dataset of documents. The score is broken down into tf and idf. The tf stands for term frequency and is the frequency of a term within a document. The equation for calculating tf is as follows:

$$tf = n_{t,d} / k_d \quad (1)$$

In this equation,  $n_{t,d}$  is the number of times the term,  $t$ , appears in the document,  $d$ , and  $k_d$  is the total number of words in the document,  $d$ .

#### 2.2.1 PMI

Pointwise Mutual Information Adjectives with same polarity tend to appear together. The affect words (adjectives, nouns, verbs and adverbs) that frequently co-occur together have the same emotional tendency. If two words co-occur more frequently, they tend to be semantically related. There are various models for measuring semantic relatedness and although they use different algorithms, they are all fundamentally based on the principle that a word's meaning can be induced by observing its statistical usage across a large sample of language. Pointwise Mutual Information (PMI) is a simple information-theoretic measure of semantic relatedness that measures the similarity between two terms by using the probability of co-occurrence. Mathematically, the PMI between two words  $x$  and  $y$  is calculated as follows:

$$PMI(x, y) = \text{co-occurrence}(x, y) / (\text{occurrence}(x) * \text{occurrence}(y)) \quad (2)$$

where  $\text{occurrence}(x)$  is the number of times that  $x$  appears in a corpus, and  $\text{co-occurrence}(x, y)$  is the number of times that  $x$  and  $y$  co-occur within a specified window  $l$  in the corpus. The corpus can

be domain-dependent or general depending on the task at hand.

### 2.2.2 Learning-based Method

Learning-based methods are being used to formulate the problem differently. Originally the problem was to determine emotions from input texts but now the problem is to classify the input texts into different emotions. Unlike keyword-based detection methods, learning-based methods try to detect emotions based on a previously trained classifier, which apply various theories of machine learning such as support vector machines and conditional random fields. To determine which emotion category should the input text belong to.

## 3. Procedure

The method we propose, consist of two approaches to detect emotions from text Word-based and Learning-based process:

### 3.1 Word-based Approach

In the Word-based approach we have used nltk (natural Language Toolkit) package, as it is best to use for human language data. In this first we assign all the Emotion Labels (joy, fear, anger, sadness, happiness) we are going to detect from the text, and also it is important to assign negation words to detect the negative emotions. After this we take the textual input and remove the unnecessary characters from the sentences, which are present there in the textual data from which detection is to be done. Tagging of the words is done to segregate the words into different category using the function `nltk.pos_tag(sentence)`, after which the stemming and the creations of data frame is done. In the stemming process all the tagged words are combined and put under a separate file and they are connected to each other, and in data frame some predefined emotion tabular form is created to make the machine learn the relation between words and labels, the tabular form includes the emotions in the first column and the words related to it included in the second & third column and the last column includes the emotion label set in which that line or paragraph will be added. And the WordNet is created between the words of the column 2 and 3. And through this the new lines whose emotion is to be detected is passed and the result is got.

### 3.2 Learning-based Approach

In the Learning-based approach we have used twitter GRAPH API to extract tweets. Extracted tweets will be saved in a excel file, there will be 2 columns, one contains the author another tweet. And we made the dataset for training and testing our machine, which will contain one more column that will be containing emotions that we have predefined. In that we divide the training dataset and testing dataset in the 3:1 ratio. After that we train the machine using the training dataset and we test it.

## 4. Results and Discussions

To deduct the emotion from the text, firstly we take the input through voice and convert it into the text using the Speech Recognition Package in python and then we try to detect emotions out of it. After the text has been converted then we use the program to detect emotion from it, the program cleans all the data means, firstly it removes all the blank spaces and not meaning full words from the sentence so that only the words which are needed to find emotion are left. Then we tokenize the words and use the word net to find the emotion from the sentence. Output is shown in the graphical as well as in coordinates form. As it ask for the input sentence from which the output has to be taken out, so we give the input and then the machine process the input which is given by the user and depict the output in the coordinated form in which the x-axis has the five emotions "Anger, Disgust, Fear, Joy, Sadness" and in the second line it shows how much the respective emotion is depicted in the input sentence.

```

C:\Users\nikharb\Desktop>python emotion_recognizer.py
What's on your mind?
Get lost!
x: ['Anger', 'Disgust', 'Fear', 'Joy', 'Sadness']
y: [0.8, 0.9, 0.0, 0.0, 0.55]

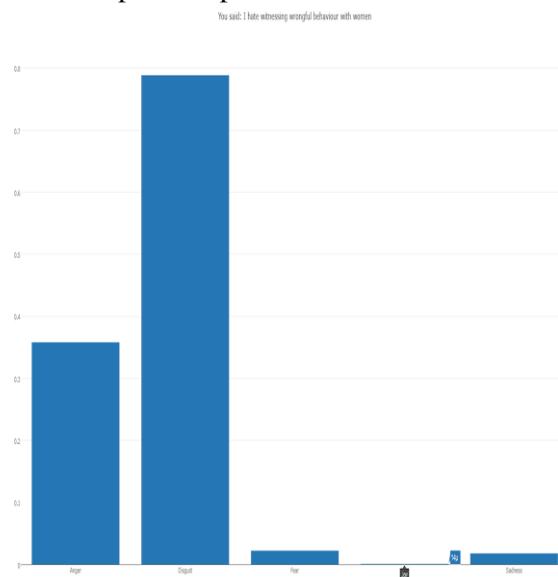
C:\Users\nikharb\Desktop>python emotion_recognizer.py
What's on your mind?
One should not underestimate the power of stupid people in groups
x: ['Anger', 'Disgust', 'Fear', 'Joy', 'Sadness']
y: [0.488939, 0.268358, 0.088436, 0.002973, 0.353572]

C:\Users\nikharb\Desktop>python emotion_recognizer.py
What's on your mind?
I respect my elders and gurus
x: ['Anger', 'Disgust', 'Fear', 'Joy', 'Sadness']
y: [0.144948, 0.161497, 0.051444, 0.375018, 0.163653]

```

The other output is in the graph form, it also depict the same output but in the graphical representation. These are the two different forms in which the output will be shown, which will help determine the emotions from the text that is entered as input by the user.

**Figure 1.** Graphical representation of Emotion from text



## 5. Conclusion

In this paper, we presented our work on text-based emotion classifications using different methods. The advantage of our system is that it is not only based on the single word in the sentence, but it also takes in to the surrounding words and then depicts the result. Moreover it considers users' experiences thanks to the historical data component. Future will consist in comforting the efficiency of the proposed textual emotion deduction modality to existing system. And also to add more emotions other than those features we have used in this paper. The best feature extraction techniques may improve the classification performance.

## References

- [1] Yasmina Douiji and Hazar Mousanifi - CARE - Intelligent Context Aware system for Recognizing Emotions from text
- [2] Agarwal A 2012 Unsupervised Emotion Detection from Text using Semantic and Syntactic Relations *IEEE/WIC/ACM International Conference on Web Intelligence and*

*Intelligent Agent Technology* pp 346-353

- [3] Haxby J V, Hoffman and Gobbini M I 2002 Human neural system for face recognition and social communication *Biol. Psychiatry* **51** pp 59-67
- [4] Nilesh 2014 Approaches of Emotion Detection from Text *International Journal of Computer Science and Information Technology Research* **2** pp 123-128\
- [5] Ekman P 2015 An Argument for Basic Emotion *International Journal of Cognition and Emotion* **6**