

# Emotion Intelligence on Text Data Repositories

Rishank Singh, Sahil Chaudhary

Department of Computer Science and Engineering

Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur,

*Email: rishank19@gmail.com, sahil.chaudhary2105@gmail.com*

Received 23 Febraury 2016, received in revised form 6 March 2016, accepted 10 March 2016

**Abstract:** Emotions are strong feeling deriving from one's circumstances, mood, or relationships with others. It involves feelings, experience, physiology, behavior, cognitions and conceptualization. Recognizing user's emotions is a major challenge for both humans and machines. Emotion detection task is very important as it can be used in domains such as software engineering, website customization, education, and gaming. Emotion detection task can be categorized into machine learning, lexical and hybrid approaches. Lexical method uses anonym and synonym of words that are directly describing basic Emotion. If any input text is applied then the word of text will be thoroughly scanned and are matched from the set of word present in the database. Machine Learning-based methods try to recognize emotions based on a previous trained classifier, which map with various machine learning classifiers to detect which emotion category, the input text belongs. In Hybrid methods, emotions are detected by using the combination of emotional keyword and learning patterns collected from training sets in addition to the information of different sciences like Human Psychology. This paper discusses the Machine Learning, lexical and Hybrid Based techniques, which are being used for emotion detection.

**Keywords:** emotion detection, Lexical-based method, Machine learning method, Hybrid method.

## 1. INTRODUCTION

Emotions and text are strongly related. Ekman's famous test for emotion recognition consists of 110 black and white pictures of Caucasian actors portraying the six universal and neutral expression which is anger, sadness, disgust, surprise, happiness and fear [1].

Recognizing behavior of human beings is a major challenge for both human and machine. It is well known that emotions are essential in human behavior that enables us to study the intellectual abilities of an individual, Technologies in computer science have not yet been able to affectively compute tasks performed by humans in order to reduce their frustrations of interactivity.

Emotions detection is a part of affective computing which is aimed to enable computer recognize emotion. Currently the system available to recognize emotions does so with the help of encoded media such as face, voice and text. The primary focus for this survey is text based analysis of human emotion.

This paper describes recent works in detecting the universal emotion plus some neutral emotion detection. It also presents the analysis of ongoing advancements in the field of semantic analysis. The methods for text based analysis can be broadly classified into lexical and machine learning approach which

will be described in the following contents. The lexical based approach is used to analyze text with the help of lexical resources such as lexicons, bag of words or ontologies etc. Whereas Machine learning approach uses classifiers and ML based algorithms based on linguistic feature. In hybrid approach based on both the approaches discussed previously. The hybrid approach is widely used as it removes the disadvantages of both machine based and lexical approach. The Paper describes the important emotion detection works done in accordance with the emotional model. The rest of paper organized into various sections as Section II describes the different approaches for emotion detection. Section III discusses Syntactic and Semantic analysis. Section IV carries the different type of Data set and finally Section V contains challenges and issues regarding the techniques.

## 2. DIFFERENT APPROACHES FOR EMOTION DETECTION

### Keyword Based method

The keyword based method is the most simple and straight forward method to detect emotional status of a person from his/her text[2]. This method uses anonyms and synonyms of the words that are directly describing Ekman's basic Emotion (i.e. happiness, anger, sadness, surprise, disgust and fear). If any input text is applied then the words of the text will be thoroughly scanned. After that the words are matched from the set of words present in our database. If any match occurs, then corresponding emotion can be analyzed[3-4]. The Problems in analyzing through keywords based method[5] are.

- *Ambivalence in Keyword Definitions*

Though detecting emotion through keyword is simple and easy but when it comes to analyze a simple word, it can have several meanings which in turns creates ambiguity. And makes the final result quite difficult.

- *Inability of Recognizing Sentences without Keywords*

Adding another disadvantage to this method is there are sentences which do not have keywords. One cannot analyze the sentence which does not contain any synonyms of our keyword.

For e.g. : Alas, his friend died in an accident.

His friend died in an accident.

The latter without “Alas” could remain undetected if “Alas” is the only keyword to detect this emotion.

• *Deficiency of Linguistic Information*

Syntax structures and semantics also have influences on expressed emotions. For example, “I laughed at him” and “He laughed at me” would suggest different emotions from the first person’s perspective. As a result, ignoring linguistic information also poses a problem to keyword-based methods. In summary, keyword-based methods should also detect not only the existence of keywords, but also their linguistic information to detect emotions more accurately.

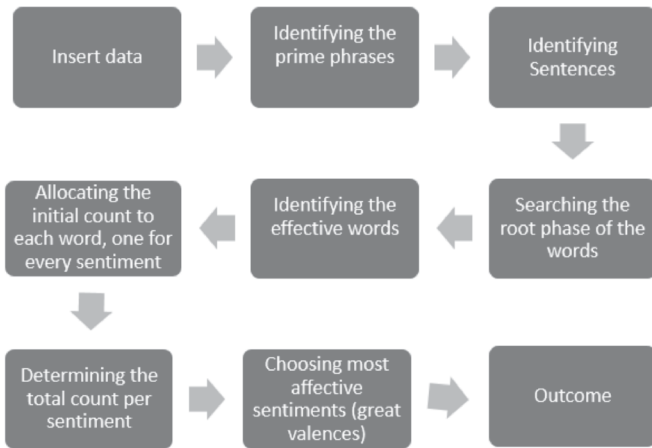


Fig 1 Keyword based method

**Learning based method**

Learning-based methods are being used to solve the problem differently. Initially the problem was to determine emotions from input texts but now the challenge is to classify the input texts into different emotions. Learning-based methods detects emotions based on a previous trained classifier/results, which is mapped with the help of various machine learning classifiers (such as support vector machines , specific statistic learning methods and decision trees), to detect which emotion category/class should the input text belongs. The major drawback of this approach is that it leads to blurred boundaries between emotion classes and a lack of context analysis. SVM is a popularly used statistical learning theory based algorithm. It has achieved empirically good performance in many fields (bioinformatics, text, image recognition). SVM assumes an even distribution between positive and negative training examples from the SVM hyper plane[6].

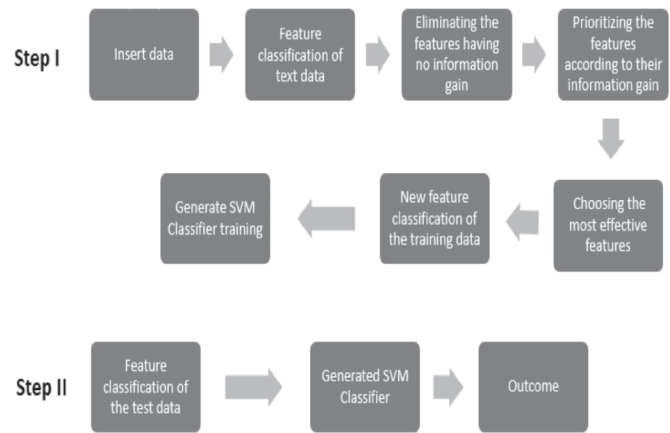


Fig 2 Learning based method

**Hybrid based method**

The hybrid system that is decomposed into keyword-based and machine learning methods[7]. If the input sentence has emotional keywords, the keyword based approach is applied. In other cases, the system uses a machine learning method, Knowledge-Based Artificial Neural Network (KBANN) to infer emotions from sentences with no emotional keywords. The keyword-based approach is based on EKD (Emotional Keyword Dictionary) that consists of words that have emotional meaning. On the other hand, KBANN network uses 3,200 sentences in the emotion-tagged corpus that come from a script of drama, novel and public web diary. This emotion recognition system combines a keyword-based approach and a KBANN machine learning approach and evaluates eight emotions (anger, fear, hope, sadness, happiness, love, thank, neutral) by separate modules[8].

**3. SYNTACTIC AND SEMANTIC ANALYSIS**

Syntactic and semantic analysis are used to test whether the text string on input is a sentence in the given (natural) language. If it is, the analysis as a result contains a brief description of the syntactic structure of the sentence that maybe in the form of a derivation tree. Such exposition are intended to make computers "understand" relationships between words (and indirectly between corresponding people, things, and actions). Syntactic analysis is applicable in large domains for example, when developing a punctuation corrector, dialogue systems with a natural language interface, or as a building block in a machine translation system[9].

As a result of all the above mentioned rules for syntactic analysis semantic and pragmatic analysis are made. It is the

most complex phase of language processing. Based on the realization about the structure of words and sentences and its meaning, the phrases, sentences and texts is analyzed, and consequently also their purpose and subsequences[10-11].

As mentioned before syntactic analysis are used to show the structure of sentence. Primarily it detects the part of speech of each word, for example noun, verb, adjective, adverb, etc. The following figure shows how any sentence in the form of an input string can be syntactically analyzed. Firstly the dependency tree for the sentence is constructed and then constraints are applied that aim to prune the tree keeping only the subtree that represents the emotion in the sentence. The reason behind the above step is to create a lexicon of emotion-related phrases that are general enough to be used to detect the emotion of any given input sentence. The output of the syntactic just mentioned is applied to an intermediate representation of the sentence that we call Emotion Recognition Rule (ERR). ERRs are composed of four types of constructs: (1) verb-noun clauses (VNCs), (2) noun clauses (NNs), (3) adjectives (JJs), and adverbs (RBs). Only these four types of constructs are chosen as they reflect actions and event descriptions[9]. Typically, emotions can be caused either by a certain action or when describing a certain incident. Figure shows a sample dependency tree extracted from the following sentence: "It was the best workshop I have ever experienced". "Experienced" is the main verb; "I" is related to "experienced"; "best" is related to "workshop" and "best workshop" is related to "experienced".

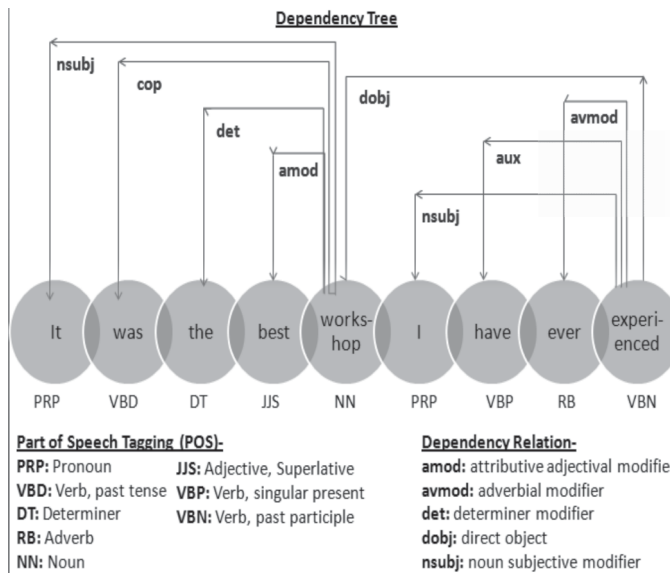


Fig 3 Dependency Tree

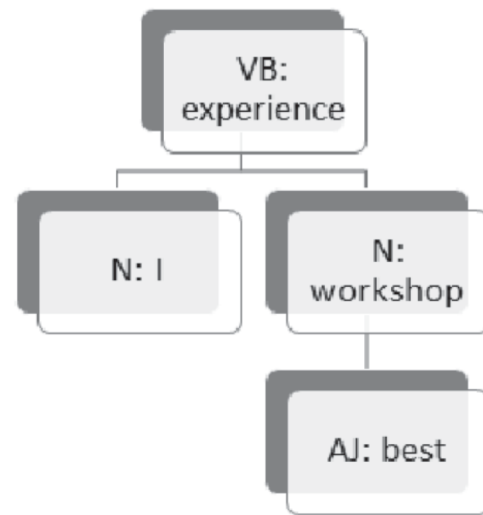


Fig 4 ERR (Emotion Recognition Rule)

#### 4. DATA SETS

The different types of emotion labeled datasets are as follows:

##### 1. Neviarouskaya et al.'s Dataset

In these datasets, ten labels were employed to explain sentences by three analyzer. These labels consist of the nine emotional categories defined by Izard (anger, disgust, fear, guilt, interest, joy, sadness, shame, and surprise) and a neutral category[1].

##### 2. Alm's Dataset

This dataset was created by Cecilla ("Cissi") Ovesdotteer include annotated sentences from fairy tales. Five emotions (happy, fearful, sad, surprised and angry-disgusted) from the Ekman's list of basic emotions were used for sentences annotations[4]. Because of data sparsity and related semantics between anger and disgust, these two emotions were merged together by the author of the dataset, to represent one class[1].

##### 3. Aman's Dataset

This dataset consists of emotion-rich sentences taken from blogs. These sentences were labelled with emotions by four analyzer. Sentences for which the annotators agreed on the emotion category are only considered[1].

##### 4. Isear dataset

The 4th type of dataset is Isear dataset. Over a period of many years during the 1990s, a large group of psychologists all over the world collected data in the ISEAR project, directed by Klaus R. Scherer and Harald Wallbott. Both psychologists and non-psychologists, were asked to report situations in which they had experienced all of 7 major emotions (joy, fear, anger, sadness, disgust, shame, and guilt) according to the response from student. In each case, the questions covered the way they had appraised the situation and how they reacted[4,12]. The final data set thus contained reports on seven emotions each by close to 3000 respondents in 37 countries on all 5 continents[13].

DATASETS	SENTENCES/EXPRESSIONS
ISEAR	When I left a man whom I believed.
FAIRY TALES	The flower could not, as on the previous evening, fold up its petals and sleep; it dropped sorrowfully.
BLOGS	I am lost in the wilderness.

Table 1: Sample sentences expressing sadness/sad from the datasets

## 5. CHALLENGES

Following are the major issues where intensive research is required.

### 1. Keyword Selection

It is very difficult to choose the right set of keywords as the sentiment is usually indicated in such a manner that it is hard to tell whether a specific term in a sentence is used in isolation or not. When correct keyword is identified, it is classified into two parts (positive or negative) which are very opposite to each other. Classification such as Topic based generally uses set of keywords to classify text in separate classes.

### 2. Sentiment is Domain Specific

Sentiment fall to a particular domain and the meaning of words is according to the circumstances that form the setting for a document or statement.

### 3. Multiple Opinions in a Sentence

One statement can contain multiple opinions when viewed with different perceptions. It is helpful to detach that clauses. It is also crucial to calculate the dominance of opinions in these clauses so that we can determine the general sentiment in the statement.

### 4. Negation Handling

Managing antithesis can be difficult in sentiment analysis. For instance, "I like this drink" and "I don't like this drink" contradict from each other by only one keyword but accordingly are to be allotted to different and opposite classes. Negation keywords are also known as polarity reversers.

### 5. Sarcasm

Usually sarcastic statements could be construed as being genuine, because they are often appropriate to the situation on a superficial level, but in fact are meant to be taken in the opposite way. For instance, "Thank you teacher for adding more assignments so that I can enjoy my weekend with more pleasure."

### 6. Comparative Sentences

Comparative keywords are used to compare differences between the two objects they modify (larger, smaller, faster, higher). Research on classifying sentences, using such keywords, as dogmatic or not is to a limit. Also the way or position of words used in comparative sentences reveal differences in the finding the correct opinion orientation. E.g. The sentence, Fruit A is tastier than Fruit B communicates a completely opposite opinion from Fruit B is tastier than fruit A..

## 7. Opinion Spam

Review sites are becoming targets for "opinion spam" -- phony positive reviews created by sellers to help sell their products, or negative reviews meant to downgrade competitors. This is called opinion spamming. In online discussion media, satisfied consumers or "innocent" parties may express specific opinions in order to further the interests of an organization in which they have an interest, such as a commercial vendor or special interest group. Web sites can also be set up for the same purpose

## 6. CONCLUSION

Emotion detection has a promising future. Major approaches towards Emotion Extraction from text have been discussed in this paper. While two approaches are promising for identifying emotion, there are still challenges to overcome. Thus we propose the emotion recognition system which combines keyword-based approach and machine learning approach i.e. a hybrid approach that relies on both, lexical analysis and machine learning analysis for classifying the six Ekman's emotional categories in user's text. For this purpose, we studied a heterogeneous dataset collected from blogs, fairy tales and new headlines, and we compared it to using each homogenous dataset separately as training data. The hybrid approach has significant improvements across all performance metrics, along with substantial reductions in misclassified cases due to inconclusive situations. Text can evoke or trigger emotions in those who read it and text can also reflect or express the emotional state of the person writing it. Thus it has large application. For instance human computer interaction.

## REFERENCES

- [1] <http://www.wikipedia.org/>
- [2] Andrius Mudinas, Dell Zhang, Mark Levene, "Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis", University of London
- [3] Mr. Nilesh M. Shelke, "Approaches of Emotion Detection from Text", International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 2, Issue 2, pp: (123-128), Month: April-June 2014
- [4] Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In Proceedings of the 2008 ACM symposium on Applied computing - SAC '08, pages 1556–1560, New York, New York, USA. ACM Press. I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [5] Sonja Gievska, Kiril Koroveshovski, Tatjana Chavdarova, "A Hybrid Approach for Emotion Detection in Support of Affective Interaction", Department of Computer Science The George Washington University, 2014 IEEE International Conference on Data Mining Workshop
- [6] Binali, Haji, Chen Wu, and Vidyasagar Potdar. "Computational approaches for emotion detection in text", 4th IEEE International Conference on Digital Ecosystems and Technologies, 2010.
- [7] Gievska, S. "A Hybrid Approach for Emotion Detection in Support of Affective Interaction" Data Mining Workshop (ICDMW), 2014 IEEE International Conference, 2014.
- [8] Soumaya Chaffar, Diana Inkpen, "Using a Heterogeneous Dataset for

- Emotion Analysis in Text”, School of Information Technology and Engineering, University of Ottawa
- [9] Ameeta Agrawal and Aijun An. 2012. Unsupervised Emotion Detection from Text Using Semantic and Syntactic Relations. In 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pages 346–353. IEEE Computer Society, December. D. P. Arnold, “Review of microscale magnetic power generation,” submitted for publication.
- [10] C. Wu, Z. Chuang, and Y. Lin, “Emotion Recognition from text Using Semantic Labels and Separable Mixture models,” ACM Transaction on Asian Language Information Processing, Vol. 5, No. 2, pp. 168-182, June 2006.
- [11] Paul Ekman. 1999. Basic emotions. In Handbook of cognition and emotion, pages 45–60. C. J. Kaufman, Rocky Mountain Research Laboratories, Boulder, CO, private communication, 2004.
- [12] Shadi Shaheen, Wassim El-Hajj, Hazem Hajj, Shady Elbassuoni, “Emotion Recognition from Text Based on Automatically Generated Rules”, 2014 IEEE International Conference on Data Mining Workshop
- [13] Sunghwan Mac Kim. 2011. Recognising Emotions and Sentiments in Text. Ph.D. thesis, University of Sydney.

