MANTECH Publications

Journal of Data Mining and Knowledge Engineering
Volume 6, Issue 1, January-June 2021

Emotion Detection from Text Using Semantic Approaches

Mohammed Zuhaib[1], Mohammed Ajmal VK[2], Alok NA[3], Favad Pasha[4], Anil Kumar KM[5]

Students[1, 2, 3, 4], Associate Professor[5]

Department of Computer Science and Engineering

JSS Science and Technology University Mysore Mysore, India

Corresponding Authors' Email id: *zuhu2195@gmail.com[1], ajmalvk24@gmail.com[2], aloknagral@gmail.com[3], favad.pasha@gmail.com[4], anilkm@sjce.ac.in[5]*

Abstract

As one of the most prosperous applications of text analysis and understanding emotions and short messaging text apperception has recently received consequential attention especially during the past several years. This is corroborated by the blogs, chats, online reviews, discussions and systematic evaluations of text analysis techniques. There are 5 major aspects for this trend, first is the commercial and social marketing applications, second is the understanding of one's feelings, third human-computer interaction , fourth text to speech generation and fifth is the natural language and machine learning approaches and technologies after 20 years of research. This model includes happy or joy, sad, disgust, anger, fear and surprise as basic emotion. In this paper emotion detection using semantic approaches and rule based approaches are described.

Keywords: *NLP (Natural Language Processing), Key-word based detection, Text Processing, Sentimental measures, Semantic approach, Parts of Speech (POS).*

INTRODUCTION

There are two main approaches for emotion detection from text, first is the keyword spotting technique and second is the Machine Learning Based Approach.

The keyword pattern matching problem can be described as the problem of finding occurrences of keywords from a given set as substrings in a given string. This problem has been studied earlier and many algorithms have been suggested for

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

solving it. In the context of emotion detection of text, this method is based on certain predefined keywords. These words are classified into categories such as happy, sad, disgust, angry, fear, surprise etc. The following are five basic steps carried out in this technique.

- Tokenization
- Identify Emotion Words
- Analysis of Intensity
- Negation Check
- Emotion

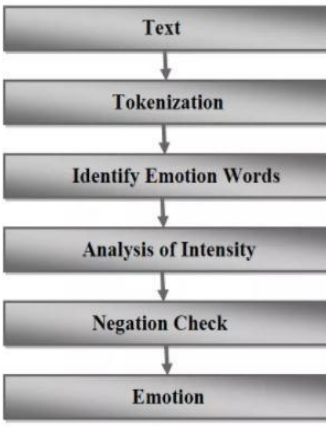


Fig 1 Keyword Spotting Technique

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 and learning-based methods try to detect emotions based on a previously trained classifier, which apply various theories of ML(Machine Learning) such as SVM(Support Vector Machines) and conditional random fields, to determine which emotion category like happy, sad, anger etc, should the input text belongs. Syntax structures and semantics also have influences on expressed emotion.

RELATED WORK

The authors of [7] present the concept and steps of using deep learning for extraction sentiments from customer reviews. The extraction pulls out the features from the customer reviews using deep learning popular methods including Naïve Bayes(NB) and Support Vector Machine(SVM) using two data sets IMDB reviews and Amazon customer reviews has been presented. This work is most similar to our work is by using deep models for sentiment analysis in customer reviews.

Tao, Jianhua[12] analyzed the functional words and classified those into emotional keyword, modifier words and metaphor words. Some context information was evaluated in order to produce a more detailed word classification. Experiments and cross testing reveal that the approach is capable of detecting emotions from text input with reasonable accuracy.

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

The authors of [6] provided the machine learning based Standford CoreNLP Part-of-Speech tagger with the Twitter model to extract essential keywords from a tweet. The testing was conducted using 6 test cases, each consisting of a human keyword generator and supervisors. As a result of the enhancements, the Turing test score of the system accuracy increased from 50% to 83.33%. Rini, Tamanna Nishat, Md Abu Naim Murad, and Md Raihan-ul-Masood.[8] paper which is most similar to our work, where they generates an Emotion Detection Model to extract emotion from text at sentence level. This method detects emotion from a text-input by searching direct emotional key words from the input. To detect emotion from text it considered six emotion class such as joy, sad, anger, disgust, fear, surprise and it produced 77% accuracies in detecting emotion from text.

[1] Armin Seyeditabari, Narges Tabari,(July 22, 2019)

Due of its applications in PS (Political Science), blogs, and other areas, In recent years, emotion identification from text has become increasingly common, human computer interaction, AI(Artificial Intelligence). In this work, we argue that current methods which are based on CML(Conventional Machine Learning) models cannot grasp the intricacy of emotional language by ignoring the sequential nature of the text, and the context.

[2] Agrawal, Ameeta, and Aijun (2012)

In this paper, we proposed a context-sensitive unsupervised approach of detecting emotions from text. Also discuss various unsupervised learning and supervised learning approach. From the experimental results, observed that the context based approach always better than the context-free approach.

[3] Minqing Hu and Bing Liu Department of Computer Science University of Illinois at Chicago

Our task is performed in three steps: (1) Mining product features that have been commented on by customers in various social media, (2) Identifying opinion sentences in each reviews and deciding whether each opinion sentence is based on positive or negative; (3) Summarize the results.

[4] TILAKRAJ, MANASA M., DEEPIKA D. SHETTY, M. NAGARATHNA, K.SHRUTHI, AND SOUGANDHIKA NARAYAN

Proposed of the system is to handle the negative sentences with positive words, it

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

decreases the sentiment score in such a manner. It also increases the sentiment scores of words like not, very, huge etc. Tweets emoticons can be classified as positive, negative and neutral feelings from the text.

Emotion from an audio speech is also analyzed. In this paper, evaluate approaches on 3 different datasets, one consisting of manual text data entry, second one being tweets in twitter and also speech by audio and give result according to their datasets that can be used.

METHODOLOGY

A. ISEAR DATASET

The International Survey on Emotion Antecedents and Reactions (ISEAR) database constructed by the Swiss NCCR (National Centre of Competence in Research) and consists of seven emotion labels like happy, sad, fear, anger, disgust, and shame obtained as a result of gathering series of data from cross-cultural questionnaire studies in 37 countries.

Three thousand participants from varying cultural backgrounds were made to fill questionnaires about their experiences and reactions toward events. The final dataset consists of 7665 sentences labelled with emotions.

B. WASSA-2017 EMOTION INTENSITIES DATASET

The Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis data was created in order to detect the intensity of emotions in tweets. It includes annotations for four distinct emotions: joy, anger, fear, and melancholy. Segregated into both training and testing sets, the data were used in the WASSA-2017 shared challenge on to help comprehend the levels of emotion intensity expressed in human languages.

C. DAILY DIALOG

The dialogues from typical human talks were used to create this dataset. It has 13118 sentences with discrete emotion labels for neutral, anger, disgust, fear, happiness, sadness, and surprise.

D. EMOTION STIMULUS DATA

Sentences conveying emotions, as well as circumstances that produce emotions are used to create the data (emotion stimulus). It has 1594 emotion sentences in it. The emotion-labelled sentences are annotated in compliance to the Ekman's basic emotion categories plus shame.

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

There are two files in the archive.

1. The Emotion Cause dataset contains eight hundred and twenty sentences with both an emotion cause and tag. The tags are in the XML format: <cause> marks the beginning of the emotion stimulus span and <\cause> marks the end of the span.

   <happy> is a good example. He is quite content these days because <he is travelling by trolley cause>

   The tags indicates the cause span and the emotion of the sentence — Happiness here.

2. The No Cause dataset in Emotion-Stimulus contains sentences marked with an emotion tag at the beginning as well as at the end, which do not carry any emotion causes. This dataset contains 1594 sentences marked with their emotion tag.
   Example: <anger>David was so angry<\anger>

   This sentence belongs to the Anger emotion class, but the cause of the emotion is not mentioned in the sentence. In both datasets, each sentence belongs to one of 7 emotions: Happy, Sad, Anger, Fear, Surprise, Disgust and Shame.

E. TEXT PROCESSING

Before applying the algorithms on the input, pre processing on the text is done. This transforms the raw input into another input format which is easy and effective for processing text. There are different methods for pre processing data such as cleaning in which it deals with punctuations, stop words, repeated letters etc. Annotation in which tokens are marked up as POS, Normalization, in which the input is structured for easy access, and extraction of valuable elements from the text, all of which are important for a specific application.

i. ***Remove punctuation :***

[18] We want interesting keywords from the given input on which processing can be done. The punctuations are uninteresting tokens in our input text structure which has to be removed. One way would be to split the input text into words by white space, then use string translation to replace all punctuations with nothing.

ii. ***Stemming :***

[19] Every word in English language has it noun, verb and adjective form. For

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

example in the given text "attract", "attracts" as noun form in POS(Parts Of Speech), "attracting" as verb form in POS, "attractive" as adjective which are having "attract" as a stem by removing '-s', '-ing', '-ive'. Now for storing all these words in database is meaningless and waste of storage in memory. Thus stemming is used for removing suffixes, prefixes and changes it to its stem word which might not be an actual word in dictionary. The benefit of stemming is that it reduces database size while also improving retrieval accuracy.

For example: exciting, excite, excited, excites is stemmed to "excit" which is not meaning full according to English dictionary.

F. IDENTIFYING SEED WORDS

First our aim was to create a base of seed words for emotions. For this we used emotion dataset from Diman Ghazi, Diana Inkpen & Stan Szpakowicz (2015).We manually identified the emotion words from the dataset and considered them to be as the base list of seed words for further analysis. We had six emotion dictionaries (keywords) for the following emotions, Anger, Fear, Happiness, Sad, Surprise and Disgust.

G. GENERATING DICTIONARIES

Then we used wordnet synonym generator to get synonyms of each word in a keyword list for each emotion into its synonyms and append it to the list if not already present in it. For all the six emotions stated above we got new enhanced list of keywords.

```
from nltk.corpus import wordnet
s=""
for disgust_word in disgust_word:
    for synset in wordnet.synsets(disgust_word):
        for lemma in synset.lemmas():
```

Fig 2 Wordnet Synonym Generator

H. POS TAGGING AND EMOTION REPRESENTATION

We used these dictionaries to understand which parts of speech better represented the different emotions. We first divided the sentences into parts of speech using nltk parts of speed tagger (nltk.pos_tag (sentence)).

For the given sentence: They have also aroused protestant, the nltk parts of speech tagger has split it into following parts of speech. ('They', 'PRP'), ('have', 'VBP'), ('also', 'RB'), ('aroused', 'VBN'), ('Protestant', 'NNP').

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

The following are some of POS (Parts of Speech) tags:

- NN: noun eg. Apple, Mango, Orange, Bat, Taj Mahal…
- NNP: Proper Noun e.g.
- PRP: Pronoun e.g. He, She, It, They, I…
- JJ: Adjective e.g. wonderful, greatest, lovely...
- RB: Adverb e.g slowly, gradually, very…

Then we identified which parts of speech better represented the above said emotions using basic linear search and recorded the results. For Anger, Sad, Disgust and Fear Adjectives comprised of majority emotion keywords. For Happy and Sad Nouns was majority keyword.

PATTERN IDENTIFICATION

Our next task was to look for specific patterns that occur in the sentences. First we analyzed bigram patterns then trigrams and followed those up to 10 word patterns. We wanted patterns that reflected negative behavior, for ex: when not appeared before happy. We created files which consisted of words that represented emotions along with 10 of the words behind that emotion including the label of that emotion.

Ex:IN for VBG fleeing RP down DT the NN statehouse NN fire NN escape TO to VB avoid NNS senators POS ' NN anger It has the part of speech that it belongs to. Using these we manually scrolled through all the sentences and iterated the process two to three times to find patterns that used to negate the sentences.

Ex: RB RB JJ



Fig 3 NLTK Tokenizer

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

This pattern was found mainly for happy emotions, if the first adverb is not or n't then it negates the meaning of the sentence. **Ex:** not totally happy

Similarly VBD RB VBN RB JJ is the four word pattern for emotion Happy. For all the emotions such patterns were identified up to 10 words.

EXPERIMENT AND RESULTS

A. ACCURACY FORMULA

Accuracy = (Total number of emotions identified correctly / Total number of sentences processed) * 100

B. SINGLE WORD FREQUENCIES OF EMOTIONS WITH RESPECT TO POS

ANGER:
Adjective: 47.61
Noun: 28.9855
Verb: 18.84
Adverb: 0.2

SAD:
Adjective: 48.521
Noun: 40.347
Verb: 12.52
Adverb: 1.391

DISGUST:
Adjective: 35.79
Noun: 35.79
Verb: 28.42

FEAR:
Adjective: 45.153
Noun: 30.49
Verb: 28.13

HAPPY:
Adjective: 73.69
Noun: 70.59 Nouns: 40.3755
Verb: 36.53 Verbs: 33.33
Adverb: 15.657

SURPRISE:
Adjective: 28.638

C. COMPARING WITHSENTIMENT ANALYSIS

Minqing Hu's and Bing Liu[3] feeling words dictionary was usedand tried to find how many sentences were discovered to be sentimental. This was done to compare see if our dictionary words were better in identifying a sentence as having emotion than the sentiment list words that identified sentiments. We applied these sentiment words on 2008 entries of training set of Daily Dailog data set and got 85.209% accuracy, meaning about 85.209 sentenced were classified as having sentiments. Similarly we applied our patterns along with dictionary words to see how many sentences were detected as having emotions and we got a score of 88.44%,meaning 88.44 percent of sentences were detected as having emotions by our model.

D. PROGRAMMING

All the patterns identified and were included in the program. We then applied the above discovered semantic rules to the Daily Dialog data set and used 1000 of its test entries. We got an accuracy of 72.4%.

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

```
        #print(list_tagged[i][j][0])
        #print(list_tagged[i][j][1])
        if(list_tagged[i][j][0].lower() in surprise_word):
            surprise_count+=1
            #print("surprise :"+list_tagged[i][j][0]+" "+lines[i])
            #emotion_write.write("surprise "+lines[i])
            if("6" in compare[i]):
                true_positive_count+=1
            #emotion_write.write("\n")
#print(surprise_count)
print(true_positive_count/len(list_tagged)*100,end='')
print(" %")
```

```
70.3187250996016 %
```

Fig 4 Accuracy using our model for train dataset

```
        #print(list_tagged[i][j][0])
        #print(list_tagged[i][j][1])
        if(list_tagged[i][j][0].lower() in surprise_word):
            surprise_count+=1
            #print("surprise :"+list_tagged[i][j][0]+" "+lines[i])
            #emotion_write.write("surprise "+lines[i])
            if("6" in compare[i]):
                true_positive_count+=1
            #emotion_write.write("\n")
#print(surprise_count)
print(true_positive_count/len(list_tagged)*100,end='')
print(" %")
```

```
72.39999999999999 %
```

Fig 5 Accuracy using our model for test dataset

E. COMAPRING WITH ML BASED MODELS

We tested the same set of data with different ML based models and got the following accuracies:

Naïve Bayes: 53%

Random Forest: 58.50%

Logistic Regression: 60.10%

SVM: 61.10%

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

F. PATTERNS IDENTIFIED

1. ANGER

a) RB JJ JJ
Not downright angry

b) VBD RB VBN
Was not offended

c) PRP VBD RB VB JJ
He did n't seem offended

2. HAPPY

a) RB RB VBN
Not particularly overjoyed

b) RB DT JJ
Not a happy

c) VBD RB VBN RB JJ
Had n't been too pleased

4. FEAR

a) VBP RB VBN
Am not frightened

b) VBD RB JJ
Was not nervous

c) VBP RB RB RB JJ
Am not now so anxious

5. SAD

a) VBD RB VBN
Were not downhearted

b) VBD RB RB JJ
Was not totally disconsolate

c) PRP VBP RB VBN
I am not downhearted

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

G. **RESULT**

SI No.	Classifier	Accuracy
1	Rule Basesd Approach	72.39
2	Naïve Bayes	53.00
3	Random Forest	58.50
4	Logistic Regression	60.10
5	SVM	61.10

CONCLUSION

Emotion Detection is the most important field of research in human-computer interaction. Researchers have put in enough effort to discern emotion from facial expressions, audio data, identifying emotions from phrases, on the other hand, is still a very young and interesting field of study. Future topics will focus on identifying sarcastic emotions. The inability to judge a sentence based on syntax and semantic information is also a difficulty due to a lack of linguistic information. In English language sentence the usage of particular syntax and semantic have different meanings at different sentences like “he laughed at me” and “I laughed at him”, Now the meaning is completely different at these two different sentences.

REFERENCES

I. Armin Seyeditabari, Narges Tabari, Shafie Gholizadeh, Wlodek Zadrozny Povoda, Lukas, Akshaj Arora, Sahitya Singh, Radim Burget, and Malay Kishore Dutta. "Emotion detection from support message." In 2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), pp. 310-313. 2015, IEEE.

II. Minqing Hu and Bing Liu Department of Computer Science University of Illinois at Chicago 851 South Morgan Street Chicago, IL 60607-7053

III. Tilakraj, Manasa M., Deepika D. Shetty, M. Nagarathna,K. Shruthi, and Sougandhika Narayan. "Emotion Finder:Detecting Emotions From Text, Tweets and Audio."

IV. Jain, Vinay Kumar, Shishir Kumar, Neha Jain, and Payal Verma. "A Multilingual Approach to Tracking Public Emotions Related to Epidemics."In 2nd International Conference and Youth School

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

Information Technology and Nanotechnology (ITNT 2016), Russia, pp. 883-889. 2016.

V. Weerasooriya, Tharindu, Nandula Perera, and S. R. Liyanage. "A method to extract essential keywords from a tweets." In Advances in ICT for Emerging Regions (ICTer), 2016 Sixteenth International Conference on, pp. 29-34. IEEE, 2016.

VI. Jain, Vinay Kumar, and Shishir Kumar. "Predictive Analysis of Emotions for Improving Customer Services." Handbook on Applying Predictive Analytics within the Service Sector (2017): 125-134.

VII. Rini, Tamanna Nishat, Md Abu Naim Murad, and Md Raihan-ul-Masood. "Emotion Detection from Text."

VIII. Agrawal, Ameeta, and Aijun An. "Unsupervised emotion detection from text using methods like semantic and syntactic relations." In Proceedings of the The 2012

IX. IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent TechnologyVolume 01, pp. 346-353. IEEE Computer Society, 2012.

X. Binali, Haji, Chen Wu, and Vidyasagar Potdar. "Computational approaches for emotion detection in text." In Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference on, pp. 172-177. IEEE, 2010.

XI. Kao, Edward Chao-Chun, Chun-Chieh Liu, TingHao Yang, Chang-Tai Hsieh, and Von-Wun Soo. "Towards text-based emotion detection a survey and possible improvements." In Information Management and Engineering, 2009. ICIME'09. International Conference on, pp. 70-74. IEEE, 2009.

XII. Tao, Jianhua. "Context based emotion detection from text input." In Eighth International Conference on Spoken Language Processing. 2004.

XIII. Balahur, Alexandra, Jesús M. Hermida, Andrés Montoyo, and Rafael Muñoz. "Emotinet: A knowledge base for emotion detection in text built on the appraisal theories." In International Conference on Application of Natural Language to Information Systems, pp. 27-39. Springer, Berlin, Heidelberg, 2011.

XIV. Calkin Suero Montero, Erkki Sutinen, and John Pajunen. Munezero, Myriam D."Are they

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved

different? Affects, feelings, emotions, sentiments etc in text." IEEE transactions on affective computing 5, no. 2 (2014): 101-111.

XV. Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. "Detection of sentiment strength in brief casual text."Journal of the Association for Information Science and Technology 61, no. 12 (2010): 2544-2558.

XVI. Wu, Chung-Hsien, Ze-Jing Chuang, and Yu-Chung Lin. "Emotion recognition from text using methods like semantic labels and separable mixture models." ACM transactions on Asian language information processing (TALIP) 5, no. 2(2006):165-183

XVII. Dhruvi D. Gosai, Himangini J. Gohil, Prof. Hardik S. Jayswal. “A Review on Text Emotion Detection and Recognition Using Natural Language Processing” ISSN 0973-4562 Volume of the International Journal of Applied Engineering Research

XVIII. https://stackoverflow.com/questions/265960/best-way-to-strip-punctuation-from-a-string-in-python.

XIX. https://pythonprogrammings.net/stemming-nltk-tutorial/

Page 46-58 © MANTECH PUBLIATIONS 2021. All Rights Reserved