A REVIEW ON A EMOTION DETECTION AND RECOGNIZATION FROM TEXT USING NATURAL LANGUAGE PROCESSING

Conference Paper · April 2018

CITATIONS
21

READS
7,466

3 authors, including:

Dhruvi Gosai
Charotar University of Science and Technology
1 PUBLICATION **21** CITATIONS
SEE PROFILE

Himangini Gohil
Charotar University of Science and Technology
1 PUBLICATION **21** CITATIONS
SEE PROFILE

A REVIEW ON A EMOTION DETECTION AND RECOGNIZATION FROM TEXT USING NATURAL LANGUAGE PROCESSING

Prof. Hardik S. Jayswal
Assistant Professor, Department of InformationTechnology,
Chandubhai S. Patel Institute of Technology,
Charusat University,
Anand, Gujarat, India
hardikjayswal.it@charusat.ac.in

Dhruvi D. Gosai
Student, Department of InformationTechnology,
Chandubhai S. Patel Institute of Technology,
Charusat University,
Anand, Gujarat, India
14it033@gmail.com

Himangini J. Gohil
Student, Department of InformationTechnology,
Chandubhai S. Patel Institute of Technology,
Charusat University,
Anand, Gujarat, India
belarajput16@gmail.com

Abstract— ***In today's world communication is done on vast area along with growth of Internet. Communication and human interaction is carried out using social sites, applications etc. in textual form. Emotion plays a vital role in human day-to-day life conveying their feelings and situation in their life. Emotions can be expressed using speech, gestures facial expression, or text. The Ekman model includes happy or joy, sad, disgust, anger, fear and surprise as basic* emotion*. Emotion Detection from textual source can be done using concepts of Natural Language Processing. In this paper emotion detection using NLP and its naive algorithm are described.***

Keywords: Emotion Recognition; Natural Language Processing; Key-word based detection; Text Processing, Sentimental measures.

I. INTRODUCTION

Emotion Detection: Sentiment Analysis is a recent field of research that is closely related to emotion detection and recognition from text. There are two types of analysis available to detect emotion. 1. Sentiment analysis 2. Emotion analysis. [37] In sentiment analysis we can detect [4] positive, negative or neutral feelings from the given user input text. In emotion analysis we can detect types of feelings such as happy, sad, anger, disgust, fear and surprise from the given user input text.

In this approach, first analyzing the given input text and then detect emotion's type of that particular given text. Detecting emotion's type from the text is quite challenging because it can't give 100% perfect result. Face expression give better result compare to the textual expression. For the human-computer interaction, detect and recognize the emotion of the given input text plays a key role. [1] Emotions may be articulated by a human's speech, face expression and written text known as speech, facial and text based emotion respectively. Most of work has been done by speech and facial emotion detection and recognition but text based emotion detection system is also very attractive for research side because now-a-days most of people write their own expression on their personal's blogs, product reviews/journals, comments etc.

[37] Emotion detection and analysis has been widely researched in psychology, finding happy planet index of the particular country and emotion marketing. By using Emotion Detection, we can generate Happy Planet Index according to level of Happiness and Healthy life style of the people among the other countries, and also give rank of the countries according to their Happy Planet Index. Emotion Detection also used in Emotion Marketing. By using Emotional Marketing we can identify that which type of people buy which type of products. Emotionally rich text can be found on product reviews, personal blogs/journals, social network websites etc.

Recognizing textual [2] emotions is a major challenge for both humans and machines, because people may not be able to recognize or state their own emotions at certain times or machines need to have proper position for emotion modeling, and also need advanced natural language processing for developing the emotion models. For this reason, we decided to develop a survey about emotion detection systems from text using natural language processing. Here in this Research paper we use natural language processing based on knowledge approach to identify the emotion type of that particular text.

II. LITERATURE SURVEY

[1] Shivhare, Shiv Naresh, and Saritha Khethawat (2012) developed a system that can be used to detect emotion from

text. In this paper, discussed about [1] Keyword spotting technique, Lexical Affinity Method, Learning-based Methods, Hybrid Methods in step by step manner. Also discuss its limitation due to detect emotion detection.

[2] Shaheen, Shadi, Wassim El-Hajj, Hazem Hajj, and Shady Elbassuoni (2014) proposed a [2] new approach for classifying emotion from textual data based on fine grained level. Performs semantic and syntactic analysis for making system context sensitive and uses WordNet and ConecptNet for covering emotion rules to generalize training set. The model shows that relation between words could lead to better accuracy than assigning value to each word. Also it performs better than EmoHeart.

[3] Tilakraj, Manasa M., Deepika D. Shetty, M. Nagarathna, K. Shruthi, and Sougandhika Narayan proposed of the system is to handles the negative sentences with positive words, it decreases the sentiment score in such a manner. It also increases the sentiment score of words like very, huge etc. Tweets emoticons can be classified as [4] positive, negative and neutral feelings from the text. Emotion from an audio is also analyzed. In this paper, evaluate approach on three different datasets, one consisting of manual text entry, second one being tweets and also speech and give result according to their dataset that can be use.

[4] TN Rini, MAN Murad, M Raihan-ul-Masood proposed three emotion detection methods such as [9] Keyword and Affect Bearing Word (ABW).Their model performs semantic analysis from semantic information, exclamatory keywords, and direct emotional keywords.

[5] Agrawal, Ameeta, and Aijun (2012) In this paper, we proposed a context-sensitive unsupervised approach of detecting emotions from text. [10] Also discuss unsupervised and supervised learning approach. From the experimental results, observed that [10] the context based approach always better than the context-free approach.

[6] Binali, Haji, Chen Wu, and Vidyasagar Potdar (2010) in this paper, established basic emotion detection models based on emotion theories and applied in computational approaches. In this paper, also define that [11] hybrid based architecture for emotion detection has been more validate and powerful with their experimental results. And also prove that [11] semantic and syntactic information can greatly improve the predication accuracy.

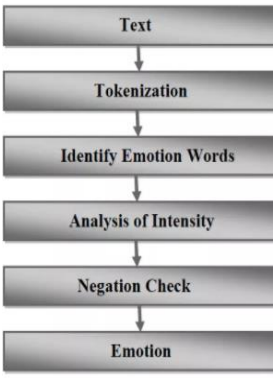
[7] Kao, Edward Chao-Chun, Chun-Chieh Liu, Ting-Hao Yang, Chang-Tai Hsieh, and Von-Wun Soo (2009) had surveyed existing research paper and review limitations as well as propose an improvement. Those improvements includes use of [12] ontology based on OCC model as knowledge base, newly-evolved vocabularies and collaborative method to detect multiple emotions in the form of case-based reasoning.

III. RELATED WORK

Emotion classification and detection is closely related to Sentimental analysis. Sentimental deals with identifying the positive, negative and neutral nature of the text whereas the Emotion Analysis deals with six basic Ekman emotions. Emotion Detection can be classified as [2] Keyword based Detection and Learning based Detection.

A. Keyword based Detection :

In this approach, emotion detection is done by extracting emotional keywords from the text. These key words are the matched with the knowledge base or the dictionary such as Thesaurus to find the emotional expression. Thesaurus is a reference work which contains synonyms, antonyms and subset of synsets. However, Keyword based Detection has drawback such as (1) unable to identify emotion from sentence that do not contain any emotional keyword, (2) ambiguity with definition of keyword as it differ from its usage and context and (3) lack of [2] linguistic information.



[1] Figure 1. Keyword Spotting Technique

B. Learning based Detection :

In this approach, the system is trained to identify the emotion. Deep Learning allows the system to understand the semantic and structure of sentence also the inter dependency of sentence. Emotion dataset is initially constructed which is tagged. This tagged dataset are then fed to neural network which trains the dataset for more accuracy and handle new data. There are various options for choosing the training model such as [6] Convolution Neural Network and Recurrent Neural Network. After training the neural network analytic reports are generated until desired accuracy is not achieved.

IV. METHODOLOGY

In this approach, we are classifying the input text into different emotions by finding the emotional content from the given English text. The emotional contents are verbs, adverbs, adjectives, phrases or combination of these keywords. For example, *"We are going on a vacation. I'm very excited"*. The keyword *"excited"* represents *"happiness"* or *"joy"*, using such keywords emotions can be classified .The source of input

to the system is textual content from social networking websites such as product reviews, comments, personal blogs, feedbacks etc. [38] The very first step is to define the structure of text in order to determine the algorithm used for emotion classification. In this approach, the structure is defined as follows:

- Each text is a list of sentences
- Each sentence is a list of tokens
- Each token is a tuple of three elements: a [38] word form (the exact word that appeared in the text), a word lemma (a generalized version of the word), and a list of associated tags

For example, *"We are going on a vacation. I'm very excited."* This input is structure as list of sentences as:

[{We are going on a vacation.} , {I'm very excited.}]

Now, each sentence in list is structured into list of tokens as:

{'We', 'are', 'going', 'on', 'a', 'vacation', '.'}, {'I', ''m', 'very', 'excited', '.'}]

Each token is further elaborated as tuple of three attributes. For example, token *"excited"* has its attributes as *('excited', 'excite', 'VB').*

A. Text Processing

Before applying the algorithms on the input, pre processing on the text is done. These transform the raw input into another format which is easy and effective for processing. There are various methods for pre processing data such as Cleaning in which it deals with punctuation, stop words, repeated letters, capitalization etc. Annotation in which the tokens are markup as POS, Normalization in which the input is organized for efficient access and extracting the useful features which is significant for particular application or task.

i. Remove punctuation :

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

ii. Repeated character :

[32] Now-a-days people on social media do not strictly follow grammar. They use different spells and shortcuts to represent their emotion as using words like *ohhhh, wowww, cooool, etc*. They will write things such as *"I likeee it"* in order to emphasize the word *"like"*. However, computers don't understand that *"likeee"* is one of the variant of *"like"* so they must be told. This method removes these annoying repeating characters in order to end up with appropriate meaningful word in English Dictionary.

iii. Negative expression replacer :

he text may contain contraction of words such as *will not* as *won't*, *cannot* as *can't*, *I am* as *I'm, I will* as *I'll*, *that is* as *that's* etc.

iv. Stop word :

The main goal of pre processing is to eliminate unwanted word which does not have any importance in application such as search queries in search engine. A stop words are these unwanted words which just occupies unnecessary space in database and increases processing time. These stop words vary from system to system. Following are some of the stop words in English language:

'some', 'against', 'at', 'can', 'these', 'ourselves', 'because', 'from', "wasn't", 'theirs', 'is', 'very', 'just', etc.

We will remove these stop words in order to save valuable space and time. They can be easily remove by storing these stop words then ignoring such stop word when encountered.

v. Stemming :

[18] Every word in English language has it noun, verb and adjective form. For example *"attract"*, has *"attracts"* as noun form, *"attracting"* as verb form, *"attractive"* as adjective which are having *"attract"* as a stem by removing *'-s', '-ing', '-ive'* etc. Now for storing all those words in database is meaningless and waste of 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 advantage of stemming is to reduce the database size and increasing the retrieval accuracy.

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

vi. Lemmatization :

[19] Lemmatization is similar to Stemming but it changes the word into its root word instead of stem word. The main difference between stemming and lemmatization is lemmatizer considers morphological analysis of word. Lemmatization is slower than stemming as it has to analyze the root word from dictionary. In order to find the correct lemma, part of speech must be specified. Words can be in the form of Noun, Adjective, Verb, and Adverb. Thus, before lemmatizing the part of speech tagging must be performed.

For example: *exciting, excite, excited, excites* is stemmed to *"excite"* in verb form.

B. Defining dictionaries of basic six expressions:

The next step is to define data dictionary which is a file in yaml format containing list of words which are classified and labeled with respective emotion tag.

Here we have defined six different dictionaries for each Ekman emotion like *happy.yml, sad.yml, fear.yml, anger.yml, disgust.yml* and *anger.yml*.

For example,

happy.yml

```
excite: [happy]
fun: [happy]
nice: [happy]
```

sad.yml

```
sorrow: [sad]
gloom: [sad]
unhappy: [sad]
```

C. Tokenization and POS tagging :

[26] In this step the input text is tokenize into tokens. According to structure defined before, each word is token if sentence is tokenized. Also each sentence is token if paragraph is tokenized. Tagging refers to classify the words based upon their parts of speech. NLTK uses word_tokenize() method for tokenization and tagging is done using pos_tag() method.

```
taged_token:  [[('We', 'We', ['PRP']), ('go'
, 'going', ['VBG']), ('vacation', 'vacation'
, ['NN'])], [('excite', 'excited', ['VBN'])]
]
```

[2] The following are some of the POS tags:

- NN: Noun e.g. Apple, Orange, Bat, Taj Mahal…
- NNP: Proper Noun e.g.
- PRP: Pronoun e.g. He, She, It, They, I…[2]
- JJ: Adjective e.g. great, best, beautiful…
- RB: Adverb e.g slowly, gradually, very…
- WP: Wh-Pronoun e.g. what, which, when, who…[2]
- CC: Conjunction e.g. and, or, but, either. or, because…[2]

D. Tagging words from dictionary.

The most important step is to find out the emotional keywords from text and classify them using re-defined dictionary. The output of this step is same as previous step but having token tagged with “happy”, “sad”, “fear”, “anger”, “disgust” or “surprise”.

```
sentiment_tag:  [[('We', 'We', ['PRP']), ('g
o', 'going', ['VBG']), ('vacation', 'vacatio
n', ['NN'])], [('excite', 'excited', ['happy
', 'VBN'])]]
```

After the tagging is performed, separation rules are applied to the output of tagging. [2] The goal of the applying these rules is to remove the non emotional content from the sentence.

Separation Rule 1: [2] Eliminate the sentence after *“but”* in input text. *‘but’* is used as connective for two ideas that contrast. The first sentence is main sentence while second sentence after ‘but’ contrast which replaces the emotion of first sentence. Thus, the sentence before *‘but’* must be ignored. The For example, *“it was a bit complicated but we had fun”.* Here the first sentence must be ignored as but contrast the emotion of sentence so *“we had fun”* is taken into consideration.

Separation Rule 2: [2] Eliminate the sentence before *“as”* if it is followed by a pronoun in input text. *‘as’* as conjunction is used when one event happens while other is in progress. The first sentence is current completed process while second sentence is continuous process .Thus, the sentence after *‘as’* must be ignored. The For example, *“We like the soup as it was served hot”.* Here the second sentence must be ignored and *“We like the soup”* is taken into consideration.

E. Calculating sentiment measure.

The sentimental measure is done by counting the frequency of happy, sad, fear, anger, disgust and surprise tags. This is the naive approach for sentiment measure.

F. Increment and decrement of sentimental measure.

[38] The strength of previous “sentimental score” using naive approach can be increased using another two data dictionary files. These can be used to increase the strength of expression which is more EKman emotion than other tags.

inc.yml

```
too: [inc]
very: [inc]
more: [inc]
```

dec.yml

```
less: [dec]
least: [dec]
quite: [dec]
```

For example: [38] the word *“good”* has more strength than *“quite good”* and less strength than *“very good”.*

G. Inverters and polarity flips

[38] Next step is to handle the polarity flips of sentence. If not handled it leads to incorrect sentimental measure. For example, *"the food at that place is not bad"*. Here word "not" is used which is used to represent negative sentence but it is use before "bad" which makes the sentence positive. Thus a new data dictionary for invert and polarity is used.

inv.yml

```
lack: [inv]
not: [inv]
```

V. EXPERIMENTAL RESULTS:

```
I am feeling glad.
HAPPY
That bus accident was too much tragic.
SAD
He looks monstrous.
DISGUST
Last night's heavy snow came as a complete surprise.
SURPRISE
The low profit figures simply confirmed my worst fears.
FEAR
The local residents were angry at the lack of parking spaces
ANGER
```

VI. FUTURE WORK AND CONCLUSION

Emotion Detection is the most [1] important field of research in human-computer interaction. A enough amount of work has been done by researchers to detect emotion from facial and audio information whereas recognizing emotions from textual data is still a fresh and hot research area.

[1] In this paper, methods which are being used to detect emotion from text are reviewed along with their limitations and new system architecture is proposed, which would perform efficiently.

REFERENCES

[1] Shivhare, Shiv Naresh, and Saritha Khethawat. "Emotion detection from text." *arXiv preprint arXiv:1205.4944* (2012).

[2] Shaheen, Shadi, Wassim El-Hajj, Hazem Hajj, and Shady Elbassuoni. "Emotion recognition from text based on automatically generated rules." In *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on*, pp. 383-392. IEEE, 2014.

[3] Jain, Ubeeka, and Amandeep Sandhu. "A Review on the Emotion Detection from Text using Machine Learning Techniques." *Emotion* (2009).

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

[5] Povoda, Lukas, Akshaj Arora, Sahitya Singh, Radim Burget, and Malay Kishore Dutta. "Emotion recognition from helpdesk messages." In *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2015 7th International Congress on*, pp. 310-313. IEEE, 2015.

[6] Jain, Vinay Kumar, Shishir Kumar, Neha Jain, and Payal Verma. "A Novel Approach to Track Public Emotions Related to Epidemics In Multilingual Data." In *2nd International Conference and Youth School Information Technology and Nanotechnology (ITNT 2016), Russia*, pp. 883-889. 2016.

[7] Weerasooriya, Tharindu, Nandula Perera, and S. R. Liyanage. "A method to extract essential keywords from a tweet using NLP tools." In *Advances in ICT for Emerging Regions (ICTer), 2016 Sixteenth International Conference on*, pp. 29-34. IEEE, 2016.

[8] 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.

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

[10] Agrawal, Ameeta, and Aijun An. "Unsupervised emotion detection from text using semantic and syntactic relations." In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01*, pp. 346-353. IEEE Computer Society, 2012.

[11] 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.

[12] Kao, Edward Chao-Chun, Chun-Chieh Liu, Ting-Hao 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.

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

[14] 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.

[15] Munezero, Myriam D., Calkin Suero Montero, Erkki Sutinen, and John Pajunen. "Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text." *IEEE transactions on affective computing* 5, no. 2 (2014): 101-111.

[16] Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. "Sentiment strength detection in short informal text." *Journal of the Association for Information Science and Technology* 61, no. 12 (2010): 2544-2558.

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

[18] https://pythonprogramming.net/stemming-nltk-tutorial/

[19] https://pythonprogramming.net/lemmatizing-nltk-tutorial/

[20] https://www.youtube.com/watch?v=t2DptEUGw6g

[21] http://www.nltk.org/

[22] https://pythonprogramming.net/tokenizing-words-sentences-nltk-tutorial/

[23] https://pypi.python.org/pypi/nltk/3.0.0

[24] http://www.nltk.org/howto/data.html

[25] http://www.nltk.org/api/nltk.tokenize.html

[26]https://kite.com/docs/python;nltk.tokenize.treebank.TreebankWordTokenizer

[27] https://www.youtube.com/watch?v=RmwKA6bhLMI

[28] https://www.powerthesaurus.org

[29] http://www.nltk.org/

[30] https://www.powerthesaurus.org/

[31] https://www.youtube.com/watch?v=9Nuc7EhdNOs

[32] https://www.youtube.com/watch?v=r37OYsdH6Z8

[33] https://www.youtube.com/watch?v=Z-y5jWZza9Q

[34] https://www.youtube.com/watch?v=uWAcgI96nEU

[35]https://pdfs.semanticscholar.org/f0e5/e9784ae89b5d3da19127f97b76e8bf1f11b4.pdf

[36] https://arxiv.org/ftp/arxiv/papers/1205/1205.4944.pdf

[37]https://www.microsoft.com/developerblog/2015/11/29/emotion-detection-and-recognition-from-text-using-deep-learning/

[38]http://fjavieralba.com/basic-sentiment-analysis-with-python.html

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