MACHINE LEARNING CA1 REPORT

on

Emotion Detection From Text
Using Machine Learning

School of Computer Science & Engineering

LOVELY PROFESSIONAL UNIVERSITY

November 2022

SUBMITED TO – Mrs. Jaspreet Kaur

Name – Vasu Goel

Registration Number- 11907868

Roll Number – RKM053A16

Table of Content

Abstract

Emotional recognition has arisen as an essential field of study that can expose a variety of valuable inputs. Emotion can be articulated in several means that can be seen, like speech and facial expressions, written text, and gestures.

Emotion can be expressed in many ways that can be seen such as facial expression and gestures, speech and by written text. Emotion Detection in text documents is essentially a content – based classification problem involving concepts from the domains of Natural Language Processing as well as Machine Learning. In this paper emotion recognition based on textual data and the techniques used in emotion detection are discussed. Emotion recognition in a text document is fundamentally a content-based classification issue, including notions from natural language processing (NLP) and deep learning fields. Hence, in this study, deep learning assisted semantic text analysis (DLSTA) has been proposed for human emotion detection using big data. Emotion detection from textual sources can be done utilizing notions of Natural Language Processing. Word embeddings are extensively utilized for several NLP tasks, like machine translation, sentiment analysis, and question answering. NLP techniques improve the performance of learning-based methods by incorporating the semantic and syntactic features of the text. The numerical outcomes demonstrate that the suggested method achieves an expressively superior quality of human emotion detection rate of 97.22% and the classification accuracy rate of 98.02% with different state-of-the-art methods and can be enhanced by other emotional word embeddings.

Keywords:- Deep Learning, text analysis, human emotion detection, NLP.

Introduction

Emotion can be conveyed in several forms, such as face and movements, voice, and written language. Emotion recognition in text documents is an issue of material – identification based on principles derived from deep learning. In day-to-day life, human emotions play an important role. Emotion can generally be understood as intuition that differs from thought or knowledge. Emotion influences an individual's personal ability to consider different circumstances and control the response to incentives. Emotional acceptance is used in many fields like medicine, law, advertising, e-learning, etc.

Detecting emotional state of a person by analyzing a text document written by him/her appear challenging but also essential many times due to the fact that most of the times textual expressions are not only direct using emotion words but also result from the interpretation of the meaning of concepts and interaction of concepts which are described in the text document. Recognizing the emotion of the text plays a key role in the human-computer interaction. Emotions may be expressed by a person's speech, face expression and written text known as speech, facial and text based emotion respectively. Sufficient amount of work has been done regarding to speech and facial emotion recognition but text based emotion recognition system still needs attraction of researchers. In computational linguistics, the detection of human emotions in text is becoming increasingly important from an applicative point of view.

The identification of feelings is one of the core aspects of object recognition in NLP. The feelings should be applied to different communication modes, including voice, facial expression, and biological signs. Text messaging is now probably the most common mode of communication. Text messages have many uses, and they are critical among texts in which emotions are efficiently understood. An insightful chat on the tweeter can understand the user's feelings and have extra sensitive and human-like responses. If a device can discern emotions from the message text, it can generate a normal speech in the text-to-speech combination.

Emotion is expressed as joy, sadness, anger, surprise, hate, fear and so on. Since there is not any standard emotion word hierarchy, focus is on the related research about emotion in cognitive psychology domain. In 2001, W. Gerrod Parrot, wrote a book named “Emotions In Social Psychology”, in which he explained the emotion system and formally classified the human emotions through an emotion hierarchy in six classes at primary level which are Love, Joy, Anger, Sadness, Fear and Surprise. Certain other words also fall in secondary and tertiary levels. Directions to improve the capabilities of current methods of text-based emotion detection are proposed in this paper.

Emotions are an important factor in detecting human activity and have multiple implementations in text messages published by users. Recovery of knowledge, contact between person and computer is useful for text analysis of human emotion. Deep learning has helped with the Semantic Text Analysis to detect human emotions through big data. Text-based source emotion tracking can be carried out using natural language processing conceptions. Word embedding is widely used for many NLP tasks, like machine translation, analysis of feelings, and question answering. NLP techniques increase academic productivity by incorporating the semantic characteristics of the text.

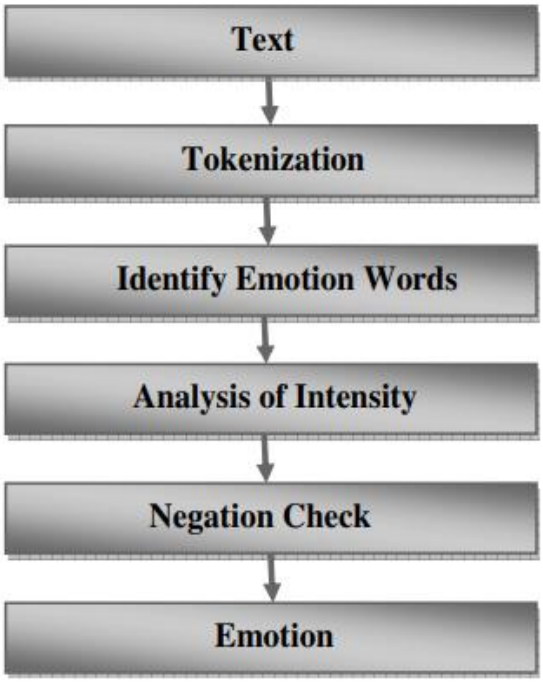
- **Dataset**

Emotion detection from text is one of the challenging problems in Natural Language Processing.The data is basically a collection of tweets annotated with the emotions behind them. We have Two columns text, sentiment. In "text" we have the raw tweet and in "sentiment", we have corresponding emotion. Task is given a raw text, predict the emotion/sentiment for it.

The data that we have is having 13 different emotions.

➢ Keyword Spotting Technique

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 in the past and algorithms have been suggested for solving it. In the context of emotion detection this method is based on certain predefined keywords. These words are classified into categories such as disgusted, sad, happy, angry, fearful, surprised etc. Process of Keyword spotting method is shown in the below figure.



Keyword spotting technique for emotion recognition consists of five steps shown in above figure where a text document is taken as input and output is generated as an emotion class. At the very first step text data is converted into tokens, from these tokens emotion words are identified and detected. Initially this technique will take some text as input and in next step we perform tokenization to the input text. Words related to emotions will be identified in the next step afterwards analysis of the intensity of emotion words will be performed. Sentence is checked whether negation is involved in it or not then finally an emotion class will be found as the required output.

➢ Lexical Affinity Method

Detecting emotions based on related keywords is an easy to use and straightforward method. Lexical Affinity approach is an extension of keyword spotting technique; it assigns a probabilistic 'affinity' for a particular emotion to arbitrary words apart from picking up emotional keywords. These probabilities are often part of linguistic corpora, but have disadvantages; firstly the assigned probabilities are biased toward corpus-specific genre of texts, secondly it misses out emotional content that resides deeper than the word-level on which this technique operates. For example the word 'accident', having been assigned a high probability of indicating a negative emotion, would not contribute correctly to the emotional assessment of phrases like 'I avoided an accident' or 'I met my girlfriend by accident'.

➢ Learning Based Methods

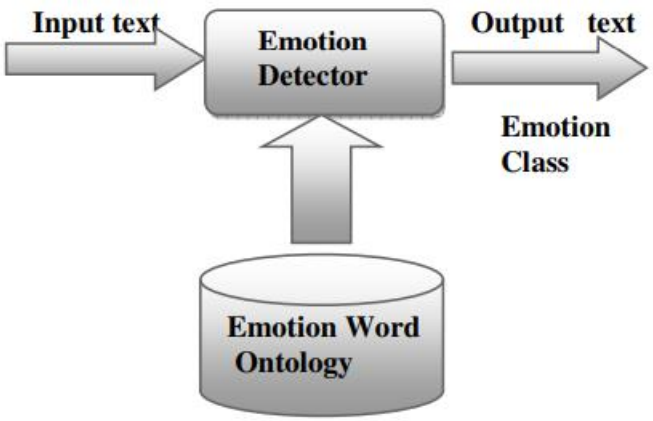
Learning-based methods are being used to formulate the problem differently. Originally the problem was to determine emotions from input texts but now the problem is to classify the input texts into different emotions. Unlike keyword-based detection methods, learning-based methods try to detect emotions 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 belongs.

➢ Hybrid Methods

Since keyword-based methods with thesaurus and naïve learning-based methods could not acquire satisfactory results, some systems use hybrid approach by combining both keyword spotting technique and learning based method, which help to improve accuracy. These semantics and attributes are associated with emotions in the form of emotion association rules. As a result, these emotion association rules, replacing original emotion keywords, serve as the training features of their learning module based on separable mixture models. This method outperforms previous approaches, but categories of emotions are still limited.

Proposed Architecture

Methods described above are modified and integrated to extend their capabilities and to improve the performance for which a simple and easy to understand model is designed shown in below diagram.



➢ Emotion Ontology

Ontology is an explicit specification of conceptualization. Ontologies have definitional aspects like high level schemas and aspects like entities and attributes; interrelationship is between entities, domain vocabulary. Ontologies provide an understanding of particular domain. Ontologies allow the domain to be communicated between persons, institutions, and application systems. Emotion word hierarchy is converted into ontology. This emotion word hierarchy is developed by W.G. parrot. Protégé, an ontology development tool is used to develop emotion ontology. Proposed ontology has class and subclass relationship format. Emotion classes at the primary level in emotion hierarchy are at the top of emotion ontology and emotion classes at the tertiary level are at the bottom of ontology. High weight age is assigned to the upper level emotion classes and low to the lower level emotion classes.

Screenshots

- Importing libraries and the datasets in the dataframe.

Emotion Detection of Text

1. Text Classification
2. Sentiment Analysis

```
In [1]: #Loading packages
        import pandas as pd
        import numpy as np
```

```
In [2]: #Loading data viz packages
        import matplotlib.pyplot as plt
        import seaborn as sns
```

```
In [3]: #Load Text cleaning
        import neattext.functions as nfx
```

```
In [4]: #Load Dataset
        df = pd.read_csv("D:\Spoonshot Assessment\intern_assignment\data.csv")
```

```
In [5]: #Preview
        df.head()
```

Out[5]:

	text	sentiment
0	@angelemotion yayyy for new videos by amazing ...	happiness
1	Surgery when I wake up	neutral
2	@ScylisSynaptic NO WAY MAN! i like 21 year olds	surprise
3	Just look @ this hellacious desk...good grief ...	hate
4	@ModelTheany tee we beefin....what was u supp...	surprise

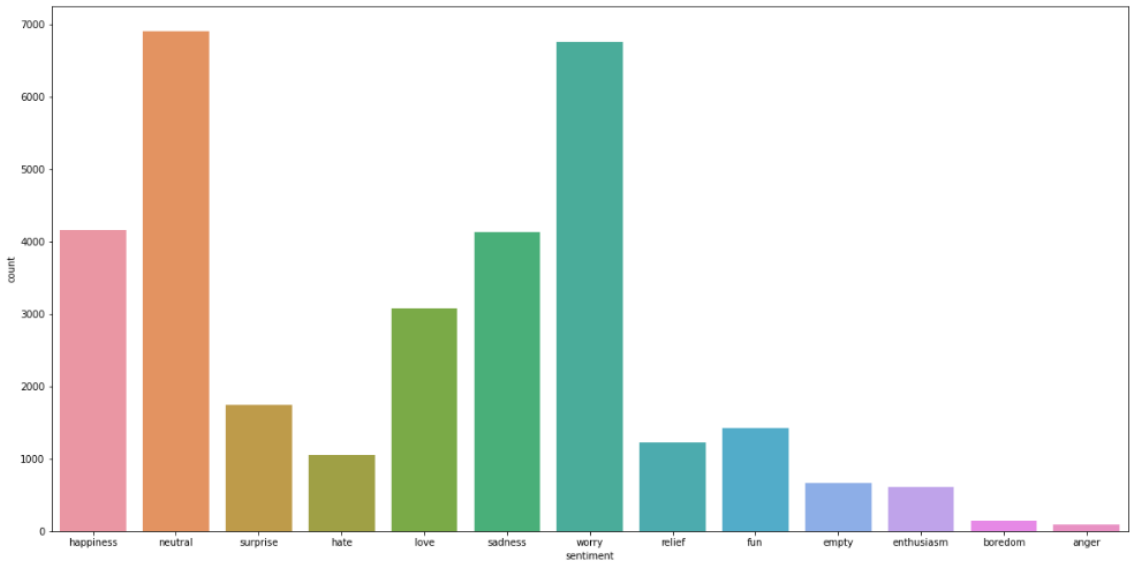
```
In [6]: #Shape
        df.shape
```

```
Out[6]: (32000, 2)
```

- Plotting emotions graph using seaborn countplot.

```
In [11]: #Using Seaborn to plot
         plt.figure(figsize=(20,10))
         sns.countplot(x='sentiment',data=df)
         plt.show
```

```
Out[11]: <function matplotlib.pyplot.show(close=None, block=None)>
```



➢ Performing Exploratory Data Analysis (EDA)

Exploratory Data Analysis(EDA)

- Text Cleaning
- Sentiment Analysis
- Keyword Extraction
  - Keywords for each emotion
  - Wordcloud

In [12]:
```
#Sentiment Analysis
from textblob import TextBlob
```

In [13]:
```
def get_sentiment(text):
    blob = TextBlob(text)
    sentiment = blob.sentiment.polarity
    if sentiment > 0:
        result = "Positive"
    elif sentiment < 0:
        result = "Negative"
    else:
        result = "Neutral"
    return result
```

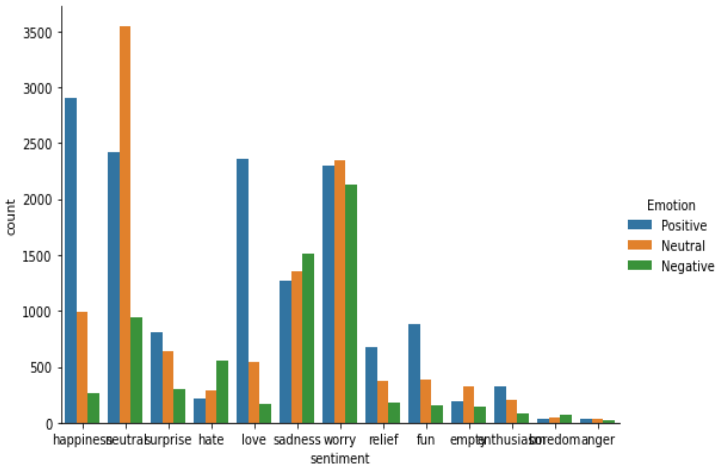
In [14]:
```
#Checking text function
get_sentiment("I am a coder")
```

Out[14]: 'Neutral'

➢ Plotting the positive, negative and neutral emotions using seaborn catplot.

In [20]:
```
sns.catplot(x='sentiment',hue='Emotion',data=df,kind='count',aspect=1.5)
```

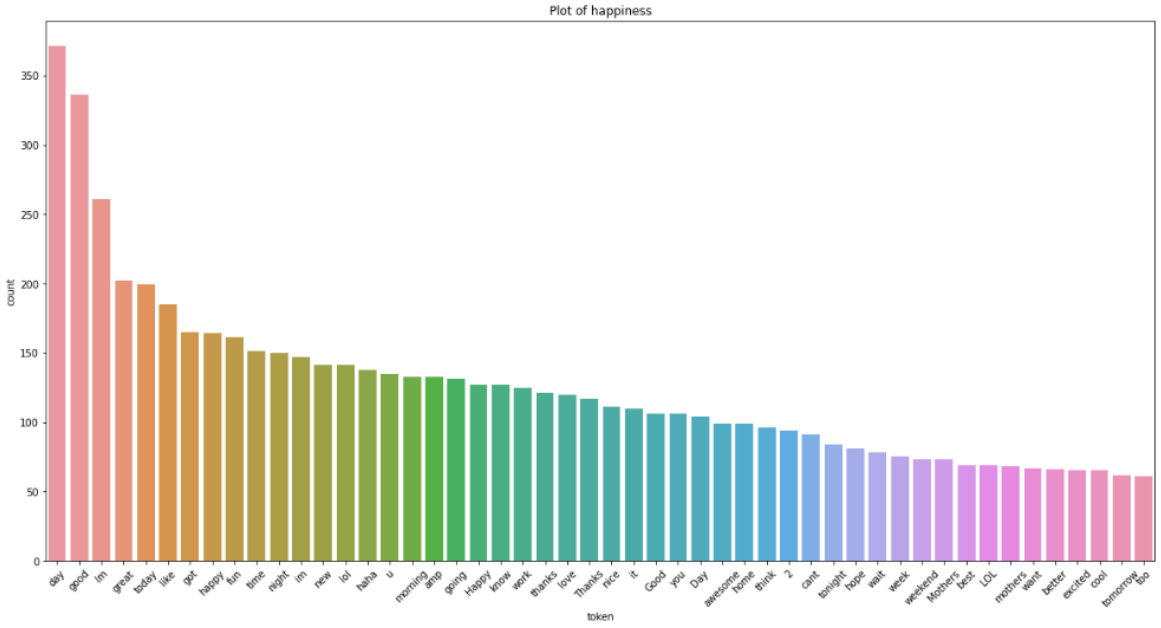
Out[20]: <seaborn.axisgrid.FacetGrid at 0x1ad20281a90>



- Plotting the most common keywords in the sentence related to happiness using seaborn barplot.

```
In [37]: #Plotting
         def plot_most_common_words(mydict,sentiment_name):
             df_01 = pd.DataFrame(mydict.items(),columns=['token','count'])
             plt.figure(figsize=(20,10))
             plt.title("Plot of {}".format(sentiment_name))
             sns.barplot(x='token',y='count',data=df_01)
             plt.xticks(rotation=45)
             plt.show()

In [38]: plot_most_common_words(keyword_happiness,"happiness")
```



- Building the model to compile and predict the accuracy.

Build our model

```
In [48]: nv_model = MultinomialNB()
         nv_model.fit(X_train,y_train)

Out[48]: MultinomialNB()

In [49]: #Accuracy
         #method
         nv_model.score(X_test,y_test)

Out[49]: 0.3104166666666667

In [50]: #Predictions
         y_pred_for_nv = nv_model.predict(X_test)

In [51]: y_pred_for_nv

Out[51]: array(['neutral', 'neutral', 'worry', ..., 'sadness', 'worry', 'worry'],
               dtype='<U10')
```

➢ Making predictions after training and building the model.

Make a single Prediction

- Vectorized our text
- Applied our model

```
In [52]: sample_text = ["I love coding"]

In [53]: vect = cv.transform(sample_text).toarray()

In [54]: #Make prediction
         nv_model.predict(vect)

Out[54]: array(['love'], dtype='<U10')

In [55]: #Check for the prediction probability(Percentage)/Confidence score
         nv_model.predict_proba(vect)

Out[55]: array([[1.11107548e-04, 1.70580346e-04, 2.98802065e-03, 2.53626923e-03,
                 2.19374656e-02, 1.00946776e-01, 3.35880841e-03, 4.25767437e-01,
                 1.63050217e-01, 1.27858477e-02, 6.54409859e-02, 2.32995676e-02,
                 1.77606917e-01]])

In [56]: #Get all classes for our model
         nv_model.classes_

Out[56]: array(['anger', 'boredom', 'empty', 'enthusiasm', 'fun', 'happiness',
                'hate', 'love', 'neutral', 'relief', 'sadness', 'surprise',
                'worry'], dtype='<U10')

In [57]: np.max(nv_model.predict_proba(vect))

Out[57]: 0.4257674367868487
```

```
In [59]: predict_emotion(sample_text,nv_model)

         Prediction:love, Prediction Score:0.4257674367868487

Out[59]: {'anger': 0.00011110754759567821,
          'boredom': 0.00017058034561305956,
          'empty': 0.0029880206544534165,
          'enthusiasm': 0.0025362692325022,
          'fun': 0.02193746564236567,
          'happiness': 0.10094677566073386,
          'hate': 0.003358808414794764,
          'love': 0.4257674367868487,
          'neutral': 0.16305021710551143,
          'relief': 0.012785847738939171,
          'sadness': 0.06544098589575412,
          'surprise': 0.023299567601199618,
          'worry': 0.17760691737368894}

In [60]: predict_emotion(["He hates running all day"],nv_model)

         Prediction:sadness, Prediction Score:0.39330603437888156

Out[60]: {'anger': 1.2231181367699192e-06,
          'boredom': 4.572376102153649e-06,
          'empty': 2.9720558876956714e-05,
          'enthusiasm': 0.00021150760690274352,
          'fun': 0.009867463420593673,
          'happiness': 0.028705379220528918,
          'hate': 0.006962268108460733,
          'love': 0.08369580367209412,
          'neutral': 0.1756159941609437,
          'relief': 0.0013185368611518112,
          'sadness': 0.39330603437888156,
          'surprise': 0.0086177075705413,
          'worry': 0.2916637889467884}
```

Conclusion

Emotion Detection can be seen as an important field of research in human-computer interaction. A sufficient 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. This paper presents DLSTA for the identification of human emotions using text analysis from big data. Textual emotion analysis can be carried out using natural language processing notions. Word embedding is commonly used for several NLP functions, including computer translation, interpretation of emotions, and question answering. The techniques of NLP enhance the efficiency of learning approaches by combining semantical and syntactic language characteristics. Emotion is conveyed in different forms, such as face and voice, gestures, and written language. Emotion can be observed with text emotion recognition, and it is a matter of information classification involving natural language processing and deep learning principles. Findings demonstrate that the suggested approach is a very promising choice for emotion recognition due to its powerful ability to learn raw data features directly. The qualitative results indicate that the proposed DLSTA approach expressly achieves the highest detection rate of 97.22 and 98.02% of classification accuracy with various emotional term embedding methods. Future work will concentrate on advancement in emotion detection, modeling the emotions' magnitude, permitting manifold emotion classes to be active concurrently, and studying alternative emotion class models.

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

References

- https://en.wikipedia.org/wiki/Machine_learning
- https://www.javatpoint.com/regression-analysis-in-machine-learning
- https://www.tutorialspoint.com/machine_learning/machine_learning_scikit_learn_algorithm.htm
- https://textblob.readthedocs.io/en/dev/
- https://towardsdatascience.com/emotion-detection-a-machine-learning-project-f7431f652b1f

Github Link:- https://github.com/Vasugoel125/Emotion-Detection-From-Text