International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Text Emotion Detection Using Machine Learning And NLP

Amal Shameem[1], Rameshbabu G[1], Vigneshwaran L[1], Sundar K[1], Mrs. K. Veena[2]

[1]UG Scholar, Department of Computer Science and Engineering, Akshaya College of Engineering and Technology, Coimbatore, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science and Engineering, Akshaya College of Engineering and Technology, Coimbatore, Tamil Nadu, India

Article Info
Volume 9, Issue 3

Page Number : 361-365

Publication Issue :
May-June-2022

Article History
Accepted : 15 May 2022
Published: 30 May 2022

ABSTRACT

In today's technological world, a majority of users across the world have access to Internet for communication via text, image, audio and video. People from diverse backgrounds exchange information on current scenarios and project their own views on them over social media. There is a need to understand and recognize the behavior of such large text information on people by analyzing their emotions. Emotions play a vital role in human interaction. We recognize emotion of a person from their speech, face gesture, body language and sign actions. Since humans use many text devices to make interactions these days, emotion extraction from the text has drawn a lot of importance. It is therefore crucial that emotions in textual conversation need to be well understood by the machines, which ultimately provide users with emotional awareness feedback. The experimental results proved that Machine learning based text emotion classification provides relatively higher accuracy compared to the existing learning methods.

Keywords: Machine Learning, Emotion Detection, NLP, Learning

I. INTRODUCTION

Social media has become an integral part of the people in 21st century. Due to rapid progress in Information & Technology sector, people have access to any kind of information at the click of a button. Moreover, with the invent of smart-phones and 4G networks, even people from the remote areas are getting connected to Tier 1 and Tier 2 cities.

With the growing population in countries like India, it has led to tremendous growth in the number of people using social networks. Social networks like Facebook, WhatsApp, Twitter, etc. has eliminated the gap between lives of people. One of the reasons to use these social networks to know the current happenings around them and to express their views and suggestions in the form of likes, share, tweets, polls, email, etc.

This has created a new category of people called netizens. Communication via social media is done in the form of text, image, audio and video which contains information and consumes space, memory and Internet bandwidth.

All these activities done on social media has resulted into vast amount of information being generated on a daily basis. Social media analysis has become a interesting field of research to understand the behavior and thoughts of people in response to social, economic, cultural, educational and all others activities happening around the world.

Social media data is in the form of unstructured data since people are from diverse backgrounds with different race, culture, language and standard of living project their ideas, views, opinions, expressions and so on the Internet.

So, it has become a challenge in recent years to extract valuable information from these ever-growing data in the form of posts, emails, blogs, micro-blogs, tweets, reviews, comments, polls and surveys on the Web about an individual, an organization or government in the process of decision-making.

These opinionated data not only exist on the Web but also within the large organizations like Google, Microsoft, Hewlett-Packard, SAP and SAS to know the opinions of their employees spread across the continents in the form of customer feedback collected from emails and call centers or results from surveys conducted by organizations. This has created strong interest for research in sentiment analysis.

II. LITERATURE REVIEW

To better understand the uses of machine learning, consider some of the instances where machine learning is applied: the self-driving Google car, cyber fraud detection, online recommendation engines—like friend suggestions on Facebook, Netflix showcasing the movies and shows you might like, and "more items to consider" and "get yourself a little something" on Amazon—are all examples of applied machine learning. All these examples echo the vital role machine learning has begun to take in today's data-rich world. Machines can aid in filtering useful pieces of information that help in major advancements, and we are already seeing how this technology is being implemented in a wide variety of industries.

2.1. Deep learning approach to text analysis for human emotion detection from big data

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 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.

2.2. AI Based Emotion Detection for Textual Big Data: Techniques and Contribution

Online Social Media (OSM) like Facebook and Twitter has emerged as a powerful tool to express via text people's opinions and feelings about the current surrounding events. Understanding the emotions at the fine-grained level of these expressed thoughts is important for system improvement. Such crucial insights cannot be completely obtained by doing AI-based big data sentiment analysis; hence, text-based emotion detection using AI in social media big data has become an upcoming area of Natural Language Processing research. It can be used in various fields such as understanding expressed emotions, human–computer interaction, data mining, online education, recommendation systems, and psychology

III. OBJECTIVE AND PROBLEM STATEMENT

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 machine learning, the detection of textual emotions is the problem of content-based classification, which is the task of natural language processing. Detecting a person's emotions is a difficult task, but detecting the emotions using text written by a person is even more difficult as a human can express his emotions in any form.

IV. EXISTING SYSTEM

The previous further discusses some recent state-of-the-art proposals in the field. discussed in relation to their major contributions, approaches employed, datasets used, results obtained, strengths, and their weaknesses. Also, emotion-labeled data sources are presented to provide neophytes with eligible text datasets for ED. Finally, the previous project presents some open issues and future research direction for text-based Emotion detection

In Previous, the existing system is based on logestic regression

- K-NN, Adaboost classifier and many more to verify. Further, machine learning algorithms
- K-NN classifier resulted the best performance with an average accuracy of 64.08%
- Adaboost classifier resulted the best performance with an average accuracy of 67.08%.

V. PROPOSED SYSTEM

This Project investigates the effectiveness Support Vector Classifier, LinearSVC, RandomForestClassifier, of mechanism for identification of textual emotions. The study was carried out on `Emotion classification' dataset with six emotional groups. In machine learning, the detection of textual emotions is the problem of content-based classification, which is the task of natural language processing. Detecting a person's emotions is a difficult task, but detecting the emotions using text written by a person is even more difficult as a human can express his emotions in any form. we propose for emotion classification in English sentences where emotions are treated as generalized concepts extracted from the sentences.

Usually, emotions are expressed as joy, sadness, anger, surprise, hate, fear, etc. Recognizing this type of emotion from a text written by a person plays an important role in applications such as chatbots, customer support forum, customer reviews etc.

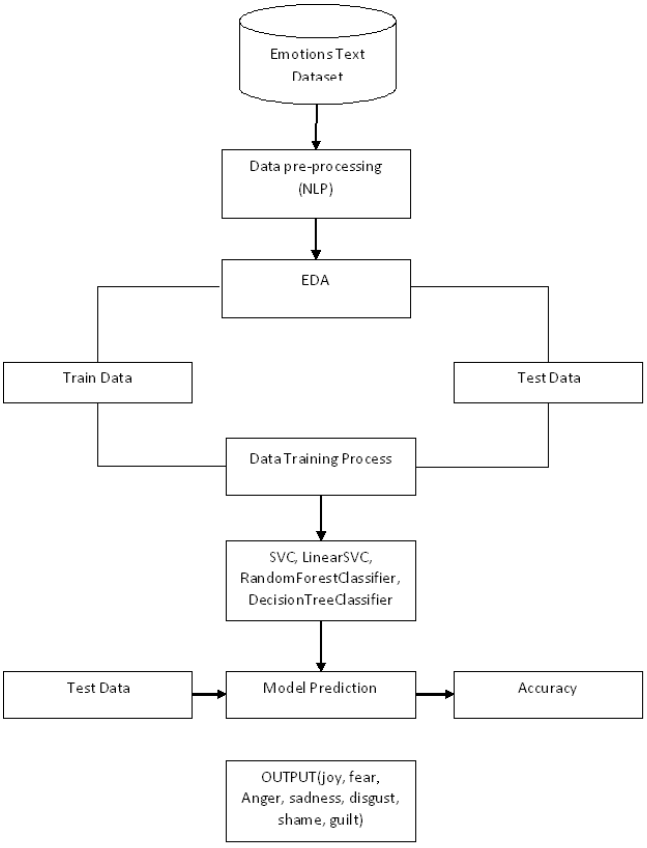

Figure 1. Data Flow Diagram

Data collection

- Our approach makes use of two annotated datasets where each sentence is annotated with one of the six emotions emotion labels. We decided to

choose this dataset as blog posts offer variety in writing styles, choice and arrangement of words and topics.

Data Pre-Processing

- Pre-processing refers to the transformations applied to our data before providing the data to the algorithm. Data Preprocessing technique is used to convert the raw data into an understandable data set. In other words, whenever the information is gathered from various sources it is collected in raw format that isn't possible for the analysis.
- Text preprocessing is **a method to clean the text data and make it ready to feed data to the model**. Text data contains noise in various forms like emotions, punctuation, text in a different case.

Feature extraction:

- DictVectorizer:
- The class DictVectorizer can be used to convert feature arrays represented as lists of standard Python dict objects to the NumPy/SciPy representation used by scikit-learn estimators.
- While not particularly fast to process, Python's dict has the advantages of being convenient to use, being sparse (absent features need not be stored) and storing feature names in addition to values.
- DictVectorizer implements what is called one-of-K or "one-hot" coding for categorical (aka nominal, discrete) features. Categorical features are "attribute-value" pairs where the value is restricted to a list of discrete of possibilities without ordering (e.g. topic identifiers, types of objects, tags, names…).

Training data and Test data

- For choosing a model we split our dataset into train and test
- Here data's are split into 3:1 ratio that means
- Training data having 70 percent and testing data having 30 percent
- In this split process preforming based on train_test_split model
- After splitting we get xtrain xtest and ytrain ytest

Model Creation

- Explore the data and choose the type of algorithm.
- Prepare and clean the dataset.
- Split the prepared dataset and perform cross validation and Perform machine learning optimisation.
- Deploy the model.

Model Prediction

Predictive modeling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data. It works by analyzing current and historical data and projecting what it learns on a model generated to forecast likely outcomes. Machine learning is a way of identifying patterns in data and using them to automatically make predictions or decisions. The emotion main methods of machine learning you will focus on are classification. Here we predict Text emtion case etc.. by algorithm performance.

Classification

In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.it is a technique where we categorize data into a given number of classes. ... Classification model: A classification model tries to draw some conclusion from the input values given for training. It will predict the class labels/categories for the new data.for classification we use machine learning algorithm .as well as prediction.

VI. CONCLUSION

Throughout this study we discussed our work on the to identify emotions based on text., SVC, Nested LinearSVC methods can be used to identify emotions in multiclass based on the results of the discussion and evaluation conducted in the previous section. RandomForestClassifier has the best accuracy among

accuracy methods. DecisionTreeClassifier has the best average performance in terms of efficiency, sensitivity and f1score at 84.7%, 74.2%, and 94.1% respectively. In this work, we introduced a new approach for classifying emotions from textual data based on a finegrained level. Our contribution lies in performing complex syntactic and semantic analysis of the sentence and using various ontologies such as Wordnet and ConceptNet in the process of emotion

VII. REFERENCES

[1]. Jeffrey F. Cohn, and Gary S. Katz. "Bimodal expression of emotion by face and voice." Proceedings of the sixth ACM international conference on Multimedia: Face/gesture recognition and their applications. ACM, 1998.

[2]. Zahra Khalili, and Mohammad Hasan Moradi. "Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG." Neural Networks, 2009. IJCNN 2009. International Joint Conference on. IEEE, 2009.

[3]. Liyanage C. De Silva, and Pei Chi Ng. "Bimodal emotion recognition." Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. IEEE, 2000.

[4]. Torao Yanaru, "An emotion processing system based on fuzzy inference and subjective observations." Artificial Neural Networks and Expert Systems, 1995. Proceedings., Second New Zealand International Two-Stream Conference on. IEEE, 1995.

[5]. EC-C. Kao, et al. "Towards Text-based Emotion Detection A Survey and Possible Improvements." Information Management and Engineering, 2009. ICIME'09. International Conference on. IEEE, 2009.

[6]. Erik Cambria, Andrew Livingstone, and Amir Hussain. "The hourglass of emotions." Cognitive behavioural systems. Springer Berlin Heidelberg, 2012. 144-157.

[7]. P. Ekman, (1999) Basic emotions. In T. Dalgleish and T. Power (Eds.) The handbook of cognition and emotion. Pp. 45-60. New York.: John Wiley & Sons.

[8]. Jeffrey T. Hancock , Christopher Landrigan, and Courtney Silver. "Expressing emotion in text-based communication." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2007.

[9]. Charles E. Osgood. Cross-cultural universals of affective meaning. University of Illinois Press, 1975.

[10]. Charles E. Osgood, and Oliver Tzeng. Language, meaning, and culture: The selected papers of CE Osgood. Praeger Publishers, 1990.

[11]. Carlo Strapparava, and Alessandro Valitutti. "WordNet Affect: an Affective Extension of WordNet." LREC. Vol. 4. 2004.

[12]. Carlo Strapparava, and Rada Mihalcea. "Learning to identify emotions in text." Proceedings of the 2008 ACM symposium on Applied computing. ACM, 2008.