VOL. 18, NO. 6, MARCH 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

EMOTION DETECTION FROM THE TEXT

K. Rajkumar[1] and P. V. Kanaka Rao[2]
[1]Department of Electronics and Communication Engineering, School of Engineering, SR University, Warangal, Telangana, India
[2]Department of Physics, University of Dodoma, Tanzania
E-Mail: rajrecw2k@gmail.com

ABSTRACT

Human emotions can be expressed in many ways through facial expressions, speech, actions, and in textual form. Emotions play an important role in determining the human's state and influence their lives in decision-making, behaviour and interaction. With the evolution of many advances in technology, researchers are able to detect the emotion of humans through facial expressions, speech, and text. Much advanced and efficient research work has been done in the field of emotion recognition through facial expression and audio, but still, there are not great works on emotion detection from text. Detecting emotions from text has wide applications in stock markets, business, decision-making, analyzing the view of people on any topic through social media conversations, tweets, blogs, and articles. Our work is focused on the detection of emotions from the text. We propose different approaches for detecting emotions through text. We implemented different approaches like the lexicon-based approach, the supervised machine learning approach with the Naïve-bayes algorithm, and the unsupervised machine learning approach with semantic similarities. After the consideration of the cons of all the approaches and the accuracies derived from our work, it is evident that the unsupervised machine learning approach with semantic similarities is the best approach with an accuracy of 78.5%.

Keywords: emotion detection, naïve bayes, semantic similarity.

1. INTRODUCTION

NLP or Natural Language Processing is a field of Artificial Intelligence that enables machines to peruse, comprehend, and get information from the dialects used by humans. It is a field that centers around the association between information science and human language. In basic terms, NLP speaks to the programmed treatment of human languages like audio or textual information.

Emotion detection from text is a trending topic in the research area. It is related to sentiment analysis. In sentiment analysis only the positive, negative, and neutral feelings from the text are detected whereas in Emotion detection the type of the feeling towards the topic is also detected i.e., emotions of the text are also recognized.

We proposed an unsupervised machine learning method for the detection of emotion. We implemented different approaches for the detection of emotion and after analyzing the disadvantages and the accuracies of those approaches. Our work is carried out on the ISAER dataset to detect the 7 basic emotions joy, surprise, disgust, anger, fear, shame and guilt. We experimented with a lexicon approach, and a supervised and unsupervised machine learning model to detect emotions. Naïve-bayes classifier of text blob is used as a supervised machine learning model and used Gensim to create an unsupervised model so that it does not require any pre-labeled training data. Gensim provides various methods to find the semantic similarity between the words thus considering all the semantic relations between the words. By the consideration of semantic relation between the words it gives a perfect model to predict the emotions with more accuracy.

2. RELATED WORK

From [1] the methodology proposed by them involves two steps- using 40% of the data to train the machine, extracting features according to the appraisal method, creation of emotion-dominant meaning hierarchies, training the classifier with the pre-processed training data, and predicting on the test data using the classifier. The architecture contains a training stage and classification stage with the training happening at the server. Dominant meaning trees are prepared from the ISEAR dataset, that tree contains all the 7 emotions joy, fear, sadness, shame, guilt, anger, and disgust. The classifier uses this information from the tree. In the proposed approach, the classifier does not use the pre-labelled examples for the classification of text, which is tie taking process. It uses the data from the constructed dominant trees. The construction of the tree is the important step, the top dominant words representing an emotion should be identified to be included in the hierarchy. These words are detected based on the frequency of the words. For a new sentence to be classified, the emotion detection algorithm calculates the values using the dominant hierarchies and returns the index of the emotion with the highest value. They followed Cohen's Kappa method to test the accuracy of the classifier. Their work proved that SVM gave better accuracies for the classification of emotion.

From [3] this paper proposes a method based on machine learning for emotion detection using supervised machine learning algorithms like support vector machine. The emotion categories used are Anger, Disgust, Sadness, Joy, and Fear. There are many approaches for recognizing emotions but every approach has its own defects and advantages, so it is preferred to use hybrid methods as they give high accuracies. The work is not based wholly on the machine learning-based approach, but it is a hybrid method combining both a keyword-based approach and a learning-based approach. Key-word based approach includes the detection of emotion based on synonyms and antonyms from the WordNet which is created manually. Their work is done by text processing followed by extraction of keywords from the dataset, checking the emotion of the keywords with the WordNet created (files of the emotion words). Each emotion has a separate file with all the words related to that particular emotion. In the learning-based approach, they extracted tweets using Graph API and the applied SVM algorithm on the training dataset. The whole work is done on the text obtained from converting a speech into text using Speech Recognition Package, then processing the text by tokenization and using the word Net to find the emotion-related. The output is shown graphically with-axis showing the five emotions Anger, Disgust, Sadness, Fear, Joy and the y-axis shows the amount of that respective emotion detected in the sentence.

The future enhancement of the work is to increase the efficiency of the proposed system and include more emotions for recognition.

From [4] this research paper focuses on the detection of emotions through unsupervised methods. There are many supervised methods, but they require large sets of annotated data which is hard to find. So, they proposed an unsupervised method that uses the semantic relationship between the words to detect the context of the words used, thus finding the emotion of the text. It does not require Wordnet to detect emotion, so it is more flexible and not dependent on a few sets of words. The methodology starts with preprocessing of text, extracting relevant affect-bearing words, and calculating the semantic relativity between the words. The semantic scores of the words are aggregated thus analyzing at the sentence level to detect the emotion of the sentence. They employed PMI for the calculation of semantic relations between the words which is a simple yet more powerful method than the previous version of LSA. Further experiments proved that stemming improved accuracy. This may be because the roots of words that belong to a specific emotion category get grouped together and this latent clustering enables more accurate frequency counts.

They applied the model extensively on various datasets: ISEAR Four emotions classification, ISEAR six emotions classification, Alm six emotions and Amag-blog dataset, thus proving the effectiveness of the proposed model.

The drawback of this model is the semantic relations are depending on the corpus of the text and we can overcome this by enhancing the derivation of semantic similarity from multiple sources.

3. PROPOSED SYSTEM

A. LEXICON APPROACH

Creation of Emotion Word Set

After the creation of the data frame and defining the emotion labels, an emotion wordset is created. Based on the emotion labels, the wordset is created. All the words of the sentences in the data set are checked with the representative words and the emotion-detecting words are stored in to the word set. This emotion wordset is used for further classification of lexicons.

Checking for Lexicons

The use of a dictionary of emotion-representing keywords is one of the primary ways to deal with emotion identification and it includes figuring out the emotion from the semantic direction of words or expressions that happen in a book. We require words speaking to all the feelings, these are made physically utilizing the feeling word set and the rundown of words speaking to a specific emotion is put away in their separate documents. By and large, this technique incorporates the portrayal of a considerable number of sentences as a lot of words. From this portrayal, the emotion representing words is recognized as dependent on the feeling word set. Mathematical functions, for example, sum and average are applied efficiently to make the last prediction with respect to the sentence's emotion. Be that as it may, it doesn't think about negation.

Classification Based on Lexicons

Check for the lexicons in the dataset by checking for the words in all sentences and creating a list of sentences and words, which act as sentence vectors and word vectors. Create a text vector by appending all the words of the word vectors. A word vector contains the words in the sentence with all the possible emotions of it as the values of the words. These words and the possible emotions are in the form of a dictionary. The combination of all the word vectors of a sentence is represented as a sentence vector. The text vector is the combination of all the sentence vectors. This text vector is used to classify the sentences and detect the emotions based on the lexicons. By applying sum and average functions to the classified words, the accuracies are calculated for every piece of text. The total accuracy is calculated by averaging all the accuracies of all the sentences.

B. SUPERVISED MACHINE LEARNINGMODEL

Splitting the Dataset

There are two stages for the working of a supervised machine learning model- the training stage and the testing stage.

The algorithm learns from the training dataset which is already pre-labeled. It contains both the input and the output contents. The algorithm learns the rules and patterns from the pre-labeled data and applies it on the test data. The aim behind the pre-labeled training set is to feed the algorithm with the "ground truth" data and learn from it.

The test data is used to know and test the algorithm and how well it has learned from the training data. The training data cannot be reused to test the algorithm as it is already aware of the output. Here is a simple flowchart that represents the process of training the algorithm and application of various functions on test and train data. We can use the train_test_split method of sklearn. preprocessing module to split the dataset or it can be done manually by the slicing data frame. The dividing

VOL. 18, NO. 6, MARCH 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ratio has a huge impact on the accuracy of the model. Usually the training data size should be greater than the testing data size such that algorithm gets more data to learn from and various special cases too, thus achieving good accuracy.

We have divided the dataset in the ratio 90-1 for good accuracy of the model. Of all the 7638 samples of our data set 6830 samples are given to the training set and the remaining 1018 samples i.e., 10% of the dataset are given to the testing set.

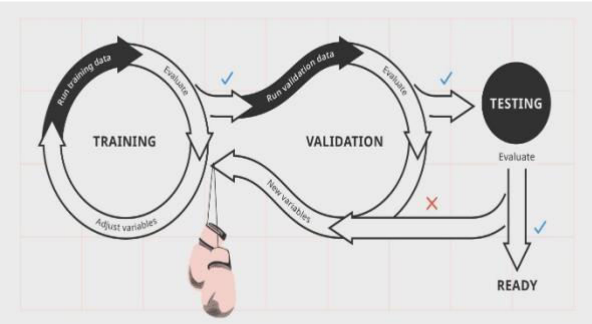

Figure-1. Splitting the dataset.

Application of Naïve Bayes Classifier

Naïve Bayes classifier is based on Bayes theorem. It works on simple probabilistic analysis in machine learning to determine the probability of the new data. The application of the algorithm is based on some assumptions. The feature of one class is independent of other feature.

The attributes of a class are independent of attributes of other classes. Naïve Bayes does the probability calculation of each attribute. This is a faster and more efficient method. Conditional probability is the probability of a class over the probability of another attribute. We can obtain the probability of an instance of the dataset, the multiplication of the conditional probabilities for each attribute for a given class. By the calculation of probabilities of the instance, the class of the instance is detected. The class with the highest probability is considered as the output.

Algorithm

$P(A/B) = P(B/A) * P(A)/P(B)$

$P(A)$ = prior probability of proposition $P(B)$ = prior probability of evidence $P(A/B)$ = posterior

$P(B/A)$ = likelihood

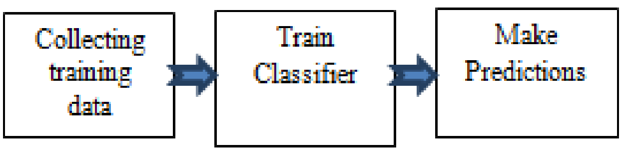

Figure-2. Working of supervised machine learning.

C. UNSUPERVISED MACHINE LEARNINGMODEL

The unsupervised Machine Learning model takes the input data but it doesn't have any corresponding output labels. In this model, there is no need to supervise the model. The goal of unsupervised learning is to model the underlying structure to learn more about the data and find unknown patterns in data by drawing inferences from the dataset. Here the machine is neither trained nor supervised by using training data but the algorithm acts on the information that is neither classified nor supervised without any guidance unlike the supervised algorithm does. They are left to their own to produce the interesting structure of the data. Since it doesn't have training data so the model categorizes based on the patterns or similarities. Unsupervised is of types i.e., Clustering and Association. Some of the algorithms that use unsupervised learning model are principal component analysis, singular value decomposition, k-means clustering etc.

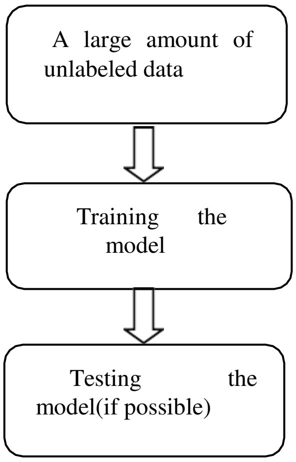

Figure-3. Working of unsupervised model.

Here we have divided the dataset into unlabeled data i.e., training and testing data in a 90:10 ratio to achieve good accuracy. And here we used the semantic similarity model in which it finds the similarity between two sentences using cosine similarity by using the Gensim package and Latent Semantic Indexing (LSI) model, here Genism, which is an NLP package that does topic modelling for humans which is used for text-processing and to work with word-vector models and LSI model is used for topic modelling. Latent Dirichlet Allocation and Latent Semantic Indexing (LSI) models are used for topic modelling.

4. RESULTS ANDANALYSIS

A. Data-frame - df

The data-frame with four columns is displayed. In the first column the type of emotions is displayed such as joy, fear, anger, sadness, disgust, shame, guilt.

In the second column the various synonyms of particular emotion are stored in tuple.

In the third column the synonyms are displayed in the form of sentence.

In the fourth column the synonyms of particular emotions are appended in the list.

VOL. 18, NO. 6, MARCH 2023 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Out[12]:

	0	1	2	3	4
0	joy	(day, feel, close, partner, other, friend, fee...	day feel close partner other friend feel peac ...	[day, feel, close, partner, other, friend, fee...	[joy]
1	fear	(time, imagin, someon, love, contact, seriou, ...	time imagin someon love contact seriou ill eve...	[time, imagin, someon, love, contact, seriou, ...	[fear]
2	anger	(had, been, obvious, unjusti, treat, had, pos...	had been obvious unjusti treat had possibl el...	[had, been, obvious, unjusti, treat, had, pos...	[anger]
3	sadness	(think, short, time, live, relat, period, life...	think short time live relat period life think ...	[think, short, time, live, relat, period, life...	[sadness]
4	disgust	(gather, found, involuntarili, sit, next, peop...	gather found involuntarili sit next peopl expr...	[gather, found, involuntarili, sit, next, peop...	[disgust]
5	shame	(realiz, wa, direct, feel, discont, partner, w...	realiz wa direct feel discont partner way wa t...	[realiz, wa, direct, feel, discont, partner, w...	[shame]
6	guilt	(feel, guilti, realiz, consid, materi, thing, ...	feel guilti realiz consid materi thing more im...	[feel, guilti, realiz, consid, materi, thing, ...	[guilt]

B. Labels

The labels are the all the seven emotion of the ISEAR dataset.

```
Out[15]: ['joy',
 'fear',
 'anger',
 'sadness',
 'disgust',
 'shame',
 'guilt',
 'joy',
 'fear',
 'anger',
 'sadness',
 'disgust',
 'shame',
 'guilt',
 'joy',
 'fear',
 'anger',
 'sadness',
 'disgust',
```

C. Prediction of Test Data

The emotion of the given test data is displayed. There are various sentences with different kinds of emotions. The emotions of all the sentences are displayed.

The input is given in the form of sentences with different kinds of emotions. The output is displayed by classifying the type of emotion.

```
data_classification("am fed up with this")
'anger'

data_classification("i love sweets")
'joy'

data_classification("i don't like to dance public")
'shame'

data_classification("i ran away from that dark place")
'fear'

data_classification("i wouldn't have beat her")
'disgust'
```

The accuracies of all the three approaches are:

a) Lexicon Approach – 44.7466%
b) Naïve Bayes Approach –62.56544%
c) Semantic Similarity Approach –78.6368%

D. Outputs

```
printing predictions of test data using semantic similarity
0 be put class leader -> shame
1 find life span china shorter west -> shame
2 girl dress foreign univers -> joy
3 be insult public -> shame
4 insult other peopl -> guilt
5 found travel best friend -> fear
6 walk home dark colleg -> sadness
7 find wa deceiv friend -> guilt
8 not do well exam -> shame
9 find best friend wa deceiv -> anger
10 fail exam did not work hard enough -> joy
11 not act promis -> disgust
12 find not ill not serious -> disgust
13 find health condit attend univers lectur -> joy
14 bought someth bad shop refus chang -> sadness
15 rel death -> sadness
16 saw peopl quarrel bu -> disgust
17 run away fire -> anger
```

5. CONCLUSIONS

The basic emotions of joy, surprise, anger, disgust, shame, and guilt are detected from the data using the ISEAR dataset for training. The work carried out by us proves that lexicon-based approach and learning-based method are not so efficient in building a model to detect emotions. Each method has its own disadvantages. Detection of emotions using an unsupervised machine learning approach proved to give more accuracy and less errors. It gave 78% accuracy in detecting the emotion of sentences given. It is due to the semantic similarity used for detection. It does not need labelled data, so it is efficient to build a model suitable for all types of datasets.

ACKNOWLEDGEMENT

We thank SR Engineering College, Warangal, Telangana, and SR University, Warangal, Telangana for supporting us during this work.

REFERENCES

[1] Razek M. A. and Frasson C. 2017. Text-Based Intelligent Learning Emotion System. Journal of Intelligent Learning Systems and Applications. 9, 17-26.

[2] V. Ramalingam1, A. Pandian2, Abhijeet Jaiswal and NikharBhatia. Emotion detection from text. National Conference on Mathematical Techniques and its Applications (NCMTA 18) IOP Publishing012027doi: 10.1088/1742-6596/1000/1/012027.

[3] Ameeta Agarwal, Aijun An. Unsupervised Emotion detection from text using semantic and syntactic

relations. Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology. Vol. 01.

[4] Simpal Kawade, Prof. K. C. Waghmare. 2017. A Survey on Identification of Emotion from Text Corpus. International Journal of Innovative Research in Computer and Communication Engineering. 5(3).

[5] Armin Seyeditabari, Narges Tabari, Wlodek Zadrozny. Emotion Detection in Text: a Review. arXiv:2004.04225[cs.CL]

[6] Lea Canales, Patricio Martínez-Barco. Emotion Detection from text: A Survey. Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC) DOI: 10.3115/v1/W14-6905

[7] Shiv Naresh Shivhare1and Prof. Saritha Khethawat2. 2012. Emotion detection from text. Data Mining and Knowledge Management Process (DKMP2012) DOI: 10.5121/csit.2012.2237.

[8] Martin D. Sykora, Thomas W. Jackson, AnnO' Brienand Suzanne Elayan. Emotive Ontology: Extracting Fine- Grained Emotions from Terse, Informal Messages. International Journal on Computer Science and Information Systems. 8(2): 106-118.

[9] Douijiyasmina, Mousannif Hajar, Al Moatassime Hassan. 2016. Using YouTube comments for text-based emotion recognition. International Conference on Ambient Systems, Networks and Technologies.

[10] M. Naveen Kumar, R. Suresh. Emotion Detection Using Lexical Chains. International Journal of Computer Applications.

[11] K. Rajkumar and P.V. Kanakarao. 2021. Pixel by Pixel Processing Technique to Characterize Tissue Imaging. ARPN Journal of Engineering and Applied Sciences. 16(5): 563-567.

[12] K. Rajkumar and Dr. V. Padmaja. 2016. Mueller Matrix Polar Decomposition of Breast Tissue. ARPN Journal of Engineering and Applied Sciences. 11(3): 1778-1781.

[13] K. Rajkumar and Dr. V. Padmaja. 2015. Polarized Laser Beam Interaction with Soft Tissue Matter. International Journal of Applied Engineering Research. 10(23): 43611-43614. ISSN: 0973-4562

[14] K. Rajkumar and Dr. V. Padmaja. 2017. Polarimetric Parameters to Categorize Normal and Malignant Thyroid Tissue. IETE Journal of Research, (Taylor & Francis), ISSN:0377-2063(Print)0974-780X(Online) 63(6): 893-897, https://doi.org/10.1080/03772063.2017.1323562