IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁒ Impact Factor 7.39 ⁒ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11722

Multilabel Text Emotion Classification

K Sathya Pramod[1], Nagasandesh N[2], Harshitha V[3], Sowndarya Spoorthi B[4], Shashank N[5]

Student, Computer Science Engineering, VVCE, Mysuru, India[1]

Professor, Computer Science Engineering, VVCE, Mysuru, India[5]

Abstract: The multi-label emotion classification task aims to identify all possible emotions in a written text that best represent the author's mental state. In recent years, multi-label emotion classification attracted the attention of researchers due to its potential applications in e-learning, health care, marketing, etc. There is a need for standard benchmark corpora to develop and evaluate multi-label emotion classification methods. The majority of benchmark corpora were developed for the English language (monolingual corpora) using tweets. The proposed work focused on English language. A multilabel emotion datasets are collected from the go emotions library. To build this project we have used both machine learning and deep learning techniques to predict the result, but compared to machine learning algorithms deep learning MLP has provided better result accuracy.

Keywords: Emotion, MLP, Multilabel, Corpora, Classification

I. INTRODUCTION

Emotions are the key to people's feelings and thoughts. Social media, such as Twitter and Facebook, have changed the language of communication. Currently, people can communicate facts, opinions, emotions, and emotion intensities on different kinds of topics in short texts. From the informative contents involved in text, not only communication information, but also private attitudes such as emotion states can be found. Human-computer interaction would be more natural and social if we can capture the emotion information occurring in texts. Researchers have tried to detect the emotion state of the intelligent interfaces users in many ways, and these works are dubbed as "affective computing" by Picard [1], which have recognized the potential and importance of affect to human computer interaction. Analyzing the emotions expressed in social media content has attracted researchers in the natural language processing research field. Emotion analysis is the task of determining the attitude towards a target or topic. The attitude can be the polarity (positive or negative) or an emotional state such as joy, anger, or sadness. Today's social media such as micro-blogs, emotions are often expressed in bilingual or multilingual text called code-switching text, and people's emotions are complex, including happiness, sadness, angry, afraid, surprise, etc. Different emotions may exist together, and the proportion of each emotion in the code-switching text is often unbalanced.

Multi-label classification problem has received more important attention nowadays. It is applicable to a wide variety of domains, such as music classification, bioinformatics, and so on. However, it has been challenging for one instance may be associated with multiple labels. In traditional single label classification methods each example has one label. But instances could have one or more labels in multi-label classification. Multi-label classification algorithms can be divided into two groups, one is the problem transformation methods (PT), and the other is the algorithm adaptation methods (AA). With the problem transformation approach, a multi-label problem is transformed into one or more single-label (i.e., binary or multi-class) problems. Specifically, single-label classifiers are learned and employed; after that, the classifiers' predictions are transformed into multi-label predictions. Different transformation methods have been proposed in the multi-label literature. The most common method is called binary relevance(BR). The idea of the binary relevance method is simple and intuitive. A multi-label problem is transformed into multiple binary problems, one problem for each label. Then, an independent binary classifier is trained to predict the relevance of one of the labels. Although binary relevance is popular in the literature, due to its simplicity, it suffers from directly modeling correlations that may exist between labels. However, it is highly resistant to overfitting label combinations, since it does not expect examples to be associated with previously-observed combinations of labels.

II. METHODOLOGY

- Collecting of datasets from tweeter
- Data pre-Processing: analysis of collected data and identifying the missing values, filling missing values and selection of attributes
- The preprocessed data is passed as input to the tfidf vectorizer to convert text data to structure to numeric data
- The tfidf vectorizer data is passed to machine learning algorithms for prediction

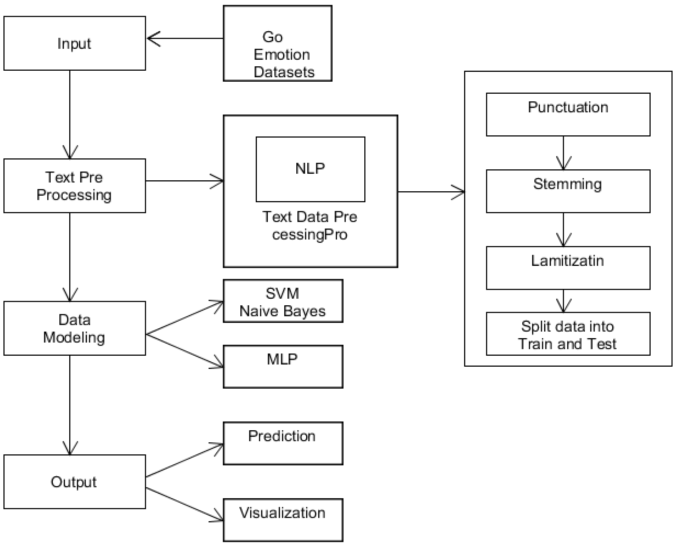
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁘ Impact Factor 7.39 ⁘ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11722

- Classification algorithms is used for prediction
- Build the model file using trained data and compare the algorithm result



Collection datasets:

- We are going to collect datasets for the prediction from the kaggle.com
- The data sets consists of 27 emotion Classes

Data Pre-Processing:

- In data pre-processing we are going to perform some image pre-processing techniques on the selected data
- And Splitting data into train and test

Data Modelling:

- The splitted train data are passed as input to the Decision tree, naïve bayes and MLP algorithm, which helps in training.
- The trained image data evaluated by passing test data to the algorithm
- Accuracy is calculatee

Build Model:

- Once the data is trained and if it showing the accuracy rate as high, then we need to build model file

SVM:

- Derivation:
- D : Set of tuples
- Each Tuple is an 'n' dimensional attribute vector
- X : (x1,x2,x3,…. xn)
- Let there be 'm' Classes : C1,C2,C3…Cm
- Naïve Bayes classifier predicts X belongs to Class Ci iff
- P (Ci/X) > P(Cj/X) for 1<= j <= m , j <> i Maximum Posteriori Hypothesis
- P(Ci/X) = P(X/Ci) P(Ci) / P(X)
- Maximize P(X/Ci) P(Ci) as P(X) is constant With many attributes, it is computationally expensive to evaluate P(X/Ci). Naïve Assumption of "class conditional independence"
- ∏= = n k P X Ci P xk Ci 1 (xk./ Ci)
- P(X/Ci) = P(x1/Ci) * P(x2/Ci) *…* P(xn/ Ci)

P(A|B) = Fraction of worlds in which B is true that also have A true
P(A ^ B) P(A|B) = ------------------ P(B)
Corollary:
P(A ^ B) = P(A|B) P(B) P(A|B)+P(¬A|B) = 1

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11722

Decision Tree Code

```
infoGain(examples, attribute, entropyOfSet)
gain = entropyOfSet
for value in attributeValues(examples, attribute):
sub = subset(examples, attribute, value)
gain -=  (number in sub)/(total number of examples) * entropy(sub)
return gain
```

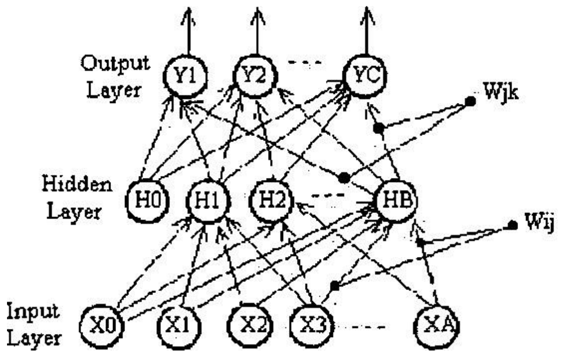
Entropy

```
entropy(examples)
'''
log2(x) = log(x)/log(2)
'''
result = 0
# handle target attributes with arbitrary labels

dictionary = summarizeExamples(examples, targetAttribute)
for key in dictionary:
proportion = dictionary[key]/total number of examples
result -= proportion * log2(proportion)
return result
```

MLP:

This work the use of direct supervised Multi-Layer Perceptron network (MLP) with one hidden layer. Its weights are adjusted by the backpropagation algorithm. In an Artificial Neural Network (ANN) the knowledge of the domain specialists is represented by the topology or the ANN and by the values of the weights used. Thus, it is considerably difficult to explain to a specialist of domain how an ANN achieved its outputs. In order to solve this problem, we utilize a rules extraction mechanism, from the trained network, of the kind IF/THEN to explain the results obtained by the network. It is worth noting that such rules are more acceptable by specialists, due to their resemblance to the human reasoning.



III. CONCLUSION

Emotions are the key to people’s feelings and thoughts. Social media, such as Twitter and Facebook, have changed the language of communication. Currently, people can communicate facts, opinions, emotions, and emotion intensities on different kinds of topics in short texts. Analysing the emotions expressed in social media content has attracted researchers in the natural language processing research field. Emotion analysis is the task of determining the attitude towards a target or topic. The attitude can be the polarity (positive or negative) or an emotional state such as joy, anger, or sadness. The proposed work focused on English language. A multilabel emotion datasets are collected from the go emotions library. To build this project we have used both machine learning and deep learning techniques to predict the result, but compared to machine learning algorithms deep learning MLP has provided better result accuracy. For Machine learning we have achieved accuracy of 60% and for Deep learning training accuracy is 96% whereas test accuracy is 86%.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 7, July 2022
DOI: 10.17148/IJARCCE.2022.11722

REFERENCES

[1]. Multi-Label Emotion Classification on Code-Mixed Text: Data and Methods
[2]. Emotion prediction of news articles from reader's perspective based on multi-label classification
[3]. Multi-label emotion classification of Urdu tweets
[4]. A Corpus-based Multi-label Emotion Classification using Maximum Entropy
[5]. A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets
[6]. Emotion classification using deep neural networks and emotional patches
[7]. A Corpus-based Multi-label Emotion Classification using Maximum Entropy
[8]. Sentiment Analysis Using Deep Learning
[9]. Emotion Detection using Deep Learning

