Sentiment and Emotion Analysis: A Novel Approach to Detect Sentiments by using Multi-Layer Perceptron

Moumita Roy[1] and Maslina Rahaman[2]
[1]Department of Computer Science, Barrackpore Rastraguru Surendranath College, India.
Email:moumita.roy.7feb@gmail.com
[2]Department of Computer Science, Barrackpore Rastraguru Surendranath College, India.
Email:maslina.thomas@gmail.com

ABSTRACT

The majority of sentiment analysis strategies rely on extensive data preprocessing, which includes carefully selecting the appropriate pattern based on the data's composition, intuitive analysis, and considerations such as vocabulary.We proposed a method for online site opinion such as IMDB movie review to analysis sentiment and emotion using the dynamic architecture for multi-layer perceptron. The proposed Multi-Layer Perceptronapproach exploits challenges associated with the distinguishing characteristics of the Movie Review facts, and the recall of fuzzy sentimental expressions. The dynamic features of perceptron of Artificial Neural Network produced a final decision feature representation consisting of multiple layers, with greater feature density.

Keywords: Sentiment, Emotion Detection, Corpus, Artificial Neural Network.

1. Introduction

The technique of evaluating a piece of data for human feelings is known as sentiment analysis. It is a way[1] to evaluate a speaker's, writer's, or other subject's attitude toward an issue, or the overall contextual polarity or emotional response to a paper, interaction, or event Analysis of public sentiment[2] is generally applied to voice of the client materials, for example, audits and overview reactions, on the web and web-based media, and medical care materials for applications that range from promoting to client support to clinical medication.

Textual opinion analysis has recently attracted a lot of interest from scientists as well as non-scientific fields like advertisement, banking, marketing, and technology.

Emotion Detection[3-4] and Sentiment Analysis is closely correlated with recognition from text, and is a relatively new area of analysis. Sentiment Analysis seeks to identify optimistic, neutral, or negative emotions in text, while Emotion Analysis seeks to identify and consider emotional types expressed in texts, such as rage, disgust, anxiety, satisfaction, sorrow, and surprise.

There are 6 emotion categories[5] that are widely used to describe humans' basic emotions, based on facial expression: rage, disgust, fear, happiness, sorrow, and surprise. "Surprise" is the most ambiguous feeling among these emotions, as it very well may be related with one or the other good or negative sentiments.

In this report, we have proposed a novel Multi-Layer Perceptron based technique for finding sentiment and emotion detection.

2. Literature Review

Our goal is to comprehend humans[6] and, all the more explicitly, human feelings or

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

human emotions. In the beyond few years,however[7], increased access to data (usually social media feeds and virtual video), inexpensive compute power, and evolving deep learning combined with natural language processing (NLP) and computer vision are enabling technologists to watch and tune in to people to investigate their opinions and feelings for analyzing their sentiments and feelings.There are a few works on neural network architectures for sentiment analysis.In (Socher et al., 2011), the authors proposed a semi-supervised method based on recursive to predict sentiments. The approach uses vector space for representation of multi-word phrases and adventures the recursive nature of sentences. The vector works on the inherent significance of the constituent, while the matrix works on word of neighborhood and the importance of how the meaning of neighboring words and phrases are changed.In (Socher et al., 2013) the authors propose the Recursive Neural Network (RNN) architecture[10],the model represents a phraseas word vectors and parse tree and then calculate sentiment by using nodes in the tree. In (Socher et al., 2011), the authors proposed a semi-supervised method that is based on recursive autoencoders for predicting sentiment. In (Chrupala, 2013), the writer proposes a simple recurrent network (SRN) [8-10]for learning continuous vector representations for sequences of characters and objects, and use them as features classifier in a conditional random field classifier to solve a character level text segmentation and labeling task.

3.Proposed Methodology

In our proposed work,Sentiment analysis follows four main processes:Extraction and Preprocessing,Feature Scaling,Splitting Training and Testing dataset, Prediction and Measure of Accuracy. And for Emotion analysis, we use the lexicon method for finding different types of human feelings.

The main workflow of this proposed work is presented in Fig. 1 and depicted in the following subsections.

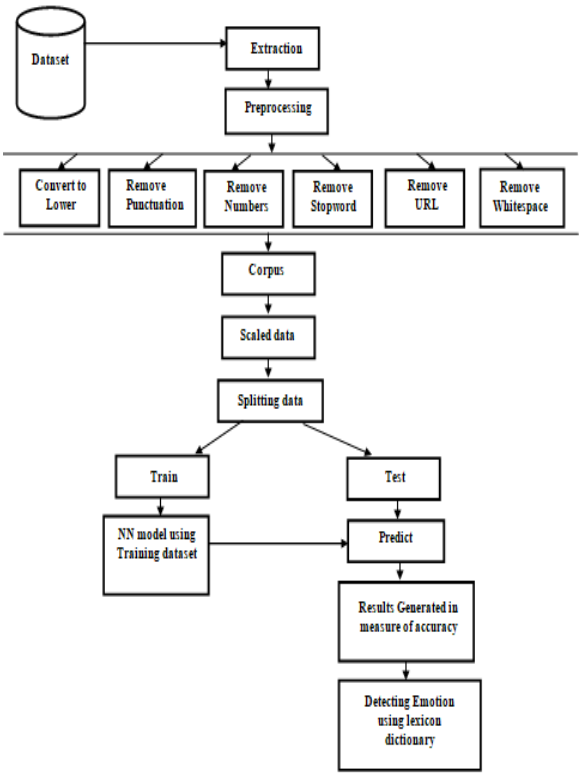

Figure 1: Architectural flow of the model

In this paper, we used movie reviews [11]as the dataset. Here, we use benchmark dataset to check whether our proposed model works properly or not.

3.1. Extraction and Preprocessing:

At first,***extraction*** is performed where the total number of distinct objects and variables and their number of occurrencesare kept in the Corpus.

After then data ***preprocessing*** is performed that involves transforming raw data into an

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

understandable format. Real-world data is regularly inadequate, incomplete, inconsistent, missing in certain behaviors or trends, and is probably going to contain many errors.[12]

Data preprocessing could be a proven technique for resolving such problems. Data preprocessing takes raw data for further processing. In our paper,the preprocessing is included conversation into lower case, removal of punctuation, numbers, unwanted URLs, whitespaces, and stop words.After preprocessing of data, the corpus contains the unique words occurring in each review of the dataset.

3.2.Featurescaling:

After that, **feature scaling** is performed. Feature scaling [13] in machine learning is one of the most important steps during the preprocessing of data before creating a machine learning model.The minimum andmaximum value of each object of the data frame is found and after getting minimum and maximum value, the feature scaling is performed-

$$\text{Scaled Value}(Reviews_{obj})=\frac{\text{Value}(Reviews_{obj})-\text{Min}(Reviews_{obj})}{\text{Max}(Reviews_{obj})-\text{Min}(Reviews_{obj})}$$

3.3. Training and Testing:

For trainingand testing purposes, the **Holdout** method is used. In this method, part of training dataset can be set aside and used as a test set, however, the common proportions fordividing thedatasets is 70% for training and 30% for testing.

The training dataset is fitted in Neural Network Model and the model is plottedwith2 hidden layers. In the initial state, 2 hidden layersareused after modification is done when it is required.

The test dataset is predicted from the learning and based on the learning it predicts the sentiment values.

3.4.Predictions and Measure of Accuracy:

In this paper,thesentiment value ofeach review in our model is divided into 5 categories such as **0**,**1**,**2**,**3**,**4** corresponding to**Somewhat Negative, Negative, Neutral, SomewhatPositive**,**Positive** respectively.

According to these sentiment values, we fitted our train dataset in themodel and then the model predicts the sentiment values of the test dataset.

After prediction, we are calculating Mean Squared Error (MSE)for getting best result-

$$\text{MSE}=\frac{1}{N}\sum_{i=0}^{N}(actual-prediction)^2$$

Where, **Σ** – a fancy symbol that means “sum”, **n** – sample size, **actual** – the actual sentiment value, and**prediction** – the predicted sentiment value.We check to see whetherMSEremains sameor updated if number of hidden layers is increased in the model. IfMSE remains the same and hence there is no change in the predicted sentiment value at each case, so there is no need to increase or decrease the numberof hiddenlayers. But if it is improved then we either increase or decrease the hidden layers till best-predicted sentiment values are achieved.

3.5.Emotion Detection:

Emotion Detection[14]is extremely complex since feeling is perplexing and not very surely known.Emotion may be deceiving when it manifests itself in a variety of ways, including our speech intonation, the content of the phrases we utter or write, our facial expressions or looks, body stance, and movements.For Emotion Analysis, we are obtaining the

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

sentiment score of each review using the lexicon approach to get sentiment score.

Eight distinct feelings are classified in our article. Rage, excitement, disgust, terror, laughter, pleasure, sorrow, and surprise are among them.

4.Results and Discussion:

Mean Squared Error obtained in different hidden layers

Number of hidden layers in NN model	MSE(mean squared error)
2	1.23989
3	0.056
4	0.09

From the above table , it is clear that if we increase the number of hidden layers in our model, the mean squared error remains decreases and hence the model with 4 hidden layers gives the best-predicted sentiment value.

Graph plot of Predicted vs. Actual Sentiment Values

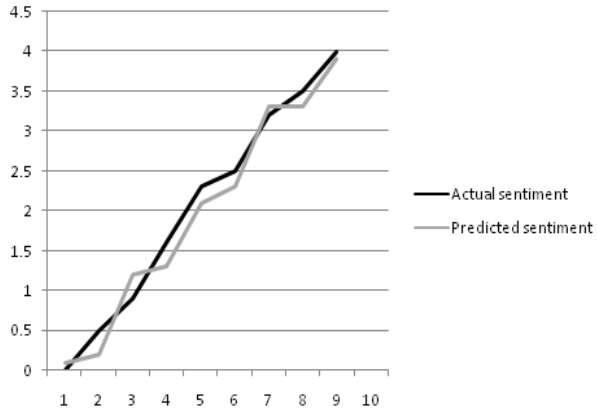

Figure2, Graph showing NN-based model predicted the sentiment values as close to the actual sentiment value.

Emotion Analysis

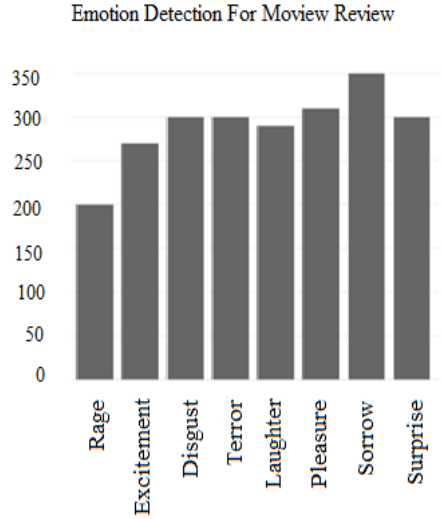

Figure 3, Graph showing different emotions based on the Movie Review dataset

Comparison chart in terms of accuracy

Methods used:	Accuracy(%)
Our model using neural network: Multilayer perceptron	82
Support vector method	80
Naïve bayes classifier	72

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

In table 2, in addition, **82%** accuracy of determining sentiment value, our model also gives the different emotions present in each data which is important as each emotion contributes to the overall sentiment of any social data.

5.Conclusion and FutureScope

The role of sentiment analysis as a classification issue has been explored in this paper.We have used movie reviews as a dataset that differ mostly in their typical length of each document, and the language style and is harder to classify**.** Experimental results show our Neural network-based system performed incredibly well, predicting sentiment values that achieved up to 82% accuracy.Finally, we can conclude that our neural network using pre-trained word vectors outperforms either of the above two machine learning methods in the challenge of Twitter sentiment analysis.

However, our model has certain limitations some of which are: it cannot detect any form of sarcasm in the statements, It cannot process any kind of emoticons used in the statements, etc. We believe that more work[15] has to be done to learn other reliable featuressuch as sarcasm, emoji, synonyms, antonyms, negations, named entities, etc. so that one can use them to improve the accuracy when running on Twitter statuses. We are looking into combining Artificial Neural Networks with other machinelearning models to gain a higher prediction accuracy.

REFERENCE :

1.https://interparestrust.org/terminology/term/sentiment%20analysis
2.https://www.questionpro.com/help/sentiment-analysis.html
3.https://deli-02.com/nwefhhf/b88daaqmfjih.php?kue=5
4.https://www.tractica.com/research/emotion-recognition-and-sentiment-analysis/
5.https://devblogs.microsoft.com/cse/2015/11/29/emotion-detection-and-recognition-from-text-using-deep-learning/
6.https://grovf.com/useCases/user-sentiment-analytics
7.http://www.iworks.am/useCases/user-sentiment-an1alytics
8.https://www.cfilt.iitb.ac.in/resources/surveys/sentiment-deeplearning-2016-prerna.pdf
9.http://nyc.lti.cs.cmu.edu/classes/11-741/Papers/santos-coling.pdf
10.https://www.researchgate.net/publication/256703371_Text_segmentation_with_character-level_text_embeddings
11.https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews/data
12.https://www.techopedia.com/definition/14650/data-preprocessing
13.https://machinelearningknowledge.ai/feature-scaling-machine-learning/
14.https://markets.businessinsider.com/news/stocks/emotion-recognition-and-sentiment-analysis-1018713190
15.https://www.studocu.com/en-us/document/university-of-alaska-fairbanks/computer-science-ii/lecture-notes/project-ml-lecture-notes-1/3161767/view

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731