Emotion Detection from Tweets and Emoticons Using Machine Learning

Prayas More[1], Abhishek Kshatriya[2], Rushikesh Kedari[3], Prof. Renuka Nagpure[4]

[1]Prayas More, Information Technology, Atharva College of Engineering
[2]Abhishek Kshatriya, Information Technology, Atharva College of Engineering
[3]Rushikesh Kedari, Information Technology, Atharva College of Engineering
[4]Prof. Renuka Nagpure, Dept. of Information Technology Engineering, Atharva College of Engineering,
Maharashtra, India

---***---

Abstract - *Many micro blogging sites have millions of people sharing their thoughts daily. We propose and investigate the sentiment from a popular real-time micro blogging service, Twitter, where real time reactions are posted by the user and we find their opinions for almost about "everything". Social networking sites like twitter, Facebook, Instagram, Orkut etc. are the great source of communication for internet users. So this becomes an important source for understanding the opinions, views or emotions of people. We extract data, i.e. tweets from Twitter in real time and apply machine learning techniques to convert them into a useful form and then use it for building sentiment classifier. Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity i.e. positive, negative. With the increase use of Internet and big explosion of text data, it has been a very significant research subject to extract valuable information from Text Ocean. To realize multi-classification for text sentiment and emoticons sentiments, this paper promotes a RNN language model based on Long Short Term Memory (LSTM). LSTM is far better than the traditional RNN. And as a language model, LSTM is applied to achieve multi-classification for text and emoticon emotional attributes.*

Key Words: RNN, LSTM, Twitter, Sentiment analysis, Opinion mining.

1. INTRODUCTION

People are posting more and more text information on Internet, and it has been a great challenge to distinguish whether the information is useful or not. As a result, it is necessary to create models to dig out valuable information, which can be used for product reviews, movie reviews, politics, sentiment analysis. However, the traditional method is very limited, as it is unable to deal with a large amount of data timely. Consequently, people are giving more importance to the efficiency of the sentimental model, and RNN is a good model. Feed-forward networks are able to take into account only a fixed context length to predict the next word, recurrent neural networks (RNN) can take advantage of all previous words. Traditional RNN language model is going further in model generalization instead of considering only the several previous words (parameter n) the recursive weights are assumed to represent short term memory. More in general we could say that RNN sees text as signal having words. Long Short-Term Memory (LSTM) neural network is different type of RNN structure. It allows to discover both long and short patterns in data and helps to eliminate the problem of vanishing gradient by training RNN.

1.1 PROPOSED SYSTEM:

The objective of this research is to create an system which will predict the emotion of user whether is it positive or negative using RNN-LSTM. LSTM are a type of RNN. Prediction is sequentially in RNN, and the hidden layer from one prediction is the hidden layer of the next prediction which will assign a memory to the network. Results from past predictions can improve future predictions. LSTM gives RNN an extra aspect that gives it fine-grained control over memory. This aspect control how much the current input matters in creating the new memory, and how much the past memories matters in creating the new memory, and what are important in generating the output. Twitter is a great source for opinions of various kinds of events and products. Detecting the sentiment of these micro blogs is a challenging task which has attracted increased research interest in recent years.

The Figure1 shows block diagram of the system. The users post their tweets in twitter. These tweets are then extracted in real time using twitter API in the form of raw data which are then saved in database. The raw dataset is converted into target dataset through Data Pre-Processing and Feature Extraction. The features of the words are selected and then machine learning techniques are applied on extracted features to classify them into its sentiment polarity that is namely positive or negative.
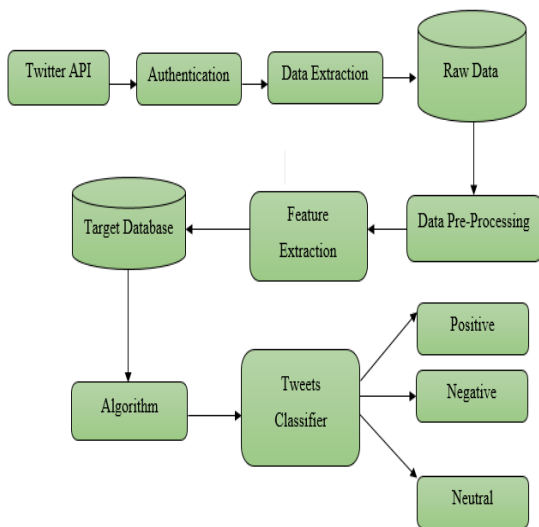
Chart -1: Proposed System

2. Data Preprocessing

Mining of Twitter data is a challenging task. The collected data is raw data. In order to apply classifier, it is essential to pre-process or clean the raw data. The pre-processing task involves removal of hashtags, uniform casing and other Twitter notations (@, RT), emoticons, URLs, stop words, decompression of slang words and compression of elongated word . [5]The following steps show the pre-processing procedure.

- To Remove Twitter notations such as hashtags (#), retweets (RT), and account Id (@).

- Remove the URLs, hyperlinks.

- Remove the stop words such as is, am, are etc. The stop words do not affect any emotions; it is just to compress the dataset.

- Compress the elongated words such as happyyy into happy.

- Decompress the slag words such as g8, f9. Mostly slang words are adjectives or nouns and they contain the highest degree level of sentiments. So it is necessary to decompress them.

- Emoji Translation: In the social media, people always use emoji to express their moods. Therefore, emoji could be useful information for sentiment analysis. We use regular expression to find out emoji pattern and translate them into positive and negative words.

2.1 Feature Extraction

The pre-processed dataset has various discrete properties. In feature extraction methods, we extract different aspects such as adjectives, verbs and nouns and later these aspects are identified as positive or negative to detect the polarity of the whole sentence. Followings are the widely used Feature Extraction methods.

- Negative Phrases: Negative words changes the meaning or orientation of the opinion. So it is evident to take negative word into account.

- Parts Of Speech (POS): Finding nouns, verbs, adjectives etc. as they are significant gauges of opinions.

2.2 Dataset for training Neural Network

Dataset contains positive and negative tweets used for training our system for prediction of user emotion.

Fig -1: Dataset

2.3 Workflow

- ENTER TWEET: User need to enter a tweet so that our system can predict whether it is positive or negative.

- ENTER SEARCH TOPIC: User need to enter a topic on which live tweets will be downloaded and our system will predict whether it is positive or negative and a pie chart will also be plotted.

3. Algorithm Used

Recurrent neural network (RNN) work as a powerful set of artificial neural network algorithm. A version of recurrent neural network works was used by DeepMind in their work playing video games with autonomous agents. Recurrent neural network work differ from feed forward neural network work because they include a feedback loop, whereby output from step x-1 is fed back to the neural network to affect the outcome of step x, and so forth for each subsequent step. For example, if a neural network is exposed to a word letter by letter, and it is asked to guess each following letter, the first letter of a word will help determine what a recurrent neural network thinks the second letter will be, etc.
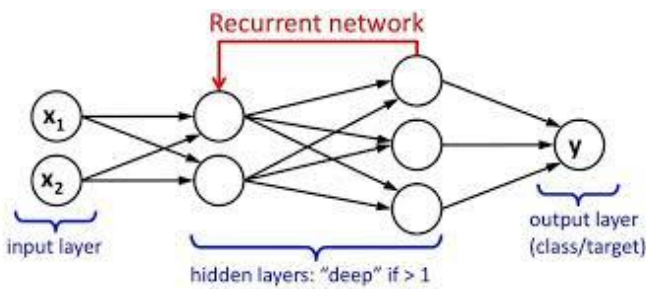
Fig -2: RNN

Long Short Term Memory neural network works – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. LSTM models are a variety of RNN. In RNN the prediction in sequence, where the hidden layer from one prediction is the hidden layer of the next prediction this will assign a memory to the neural network work, therefore, results 'from earlier estimation could lead to improve future predictions. LSTM gives RNN more features to a extreme control over memory; this aspects control how much the present input matters for forming the new memory, also how much the past memories matters in creating the new memory, and what parts of the memory are essential is producing the output.
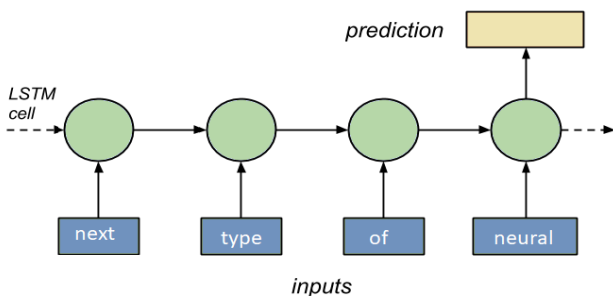

Fig -3: LSTM

4. Results

This section deals with the results of the system implemented and are discussed below:

- The given figure below shows the Enter text window in which user need to enter tweet to find its emotion i.e. whether it is positive or negative
- Then user can enter a topic on which live tweets will be downloaded and our system will predict its polarity.
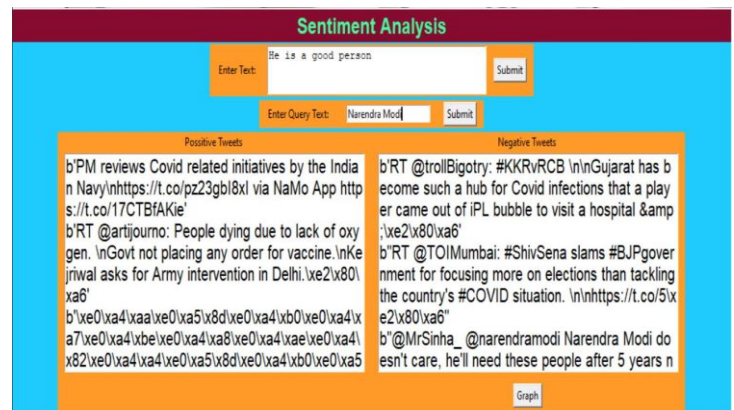

Fig -4: GUI Interface

- Also when the result is predicted , results are saved in file with its sentiment and a graph
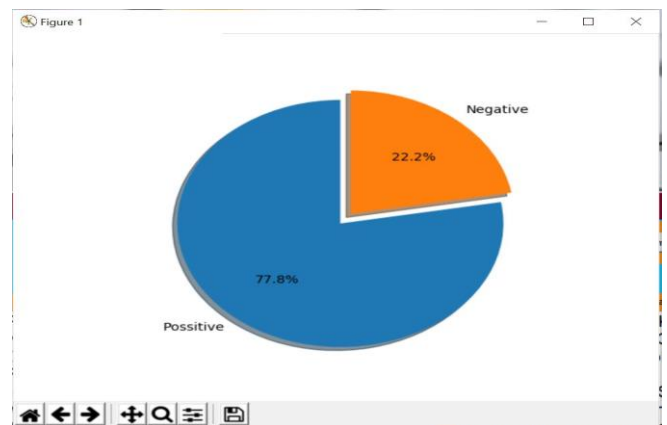

Fig -5: Pie Chart

5. Conclusions

In this paper, an improved RNN language model is put forward using LSTM, which successfully covers all history sequence information and performs better than conventional RNN. It is applied to achieve multi-classification for text emotional attributes, and identifies text emotional attributes more accurately. Even Today Sentiment Analysis is still a difficult and Complex problem in computer Science. Sentiments are express by Humans in Different Ways. The objective of this work is to create a system for finding the emotion from tweets and emoticons for users in a user friendly way.

This system has lot of future scope as the need for this system is felt very greatly. The few numbers of future implementations are:

In future work, we could try these improvement programs or use different models combination to improve the performance of text sentiment analysis. The models will be trained and validated against a test dataset. We apply machine learning techniques to solve twitter sentiment analysis problem.

Also, Twitter sentiment analysis comes under the category of text and opinion mining. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don't perform well when the number of classes are increased. Also it's still not tested that how accurate the model will be for other topics. Hence sentiment analysis has a very bright scope of development in future.

6. REFERENCES

[1] Antonio Lopardo, Marco Brambilla "Analyzing and Predicting the US Midterm Elections on Twitter with Recurrent Neural Networks" 2018 IEEE International Conference on Big Data (Big Data)

[2] Rohith.V, D. Malathi "Sentiment Analysis On Twitter: A Survey" International Journal of Pure and Applied Mathematics Volume 118 No. 22 2018, 365-375

[3] Fenna Miedema "Sentiment Analysis with Long Short-Term Memory networks" Research Paper Business Analytics Vrije Universiteit Amsterdam August 1, 2018

[4] Abdalraouf Hassan "Sentiment Analysis With Recurrent Neural Network And Unsupervised Neural Language Model " March 2017

[5] Mitali Desai, Mayuri A. Mehta, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey" IEEE System, International Conference on Computing, Communication and Automation (ICCCA2016)

BIOGRAPHIES

Rushikesh Kedari is currently a final year student who is pursuing Bachelor of Engineering degree in Information Technology from Mumbai University. His area of research includes UI/UX design, Web development and data science.

Prayas More is currently a final year student who is pursuing Bachelor of Engineering degree in Information Technology from Mumbai University. His area of research includes Data Science and Machine Learning.

Abhishek Kshatriya is is currently a final year student who is pursuing Bachelor of Engineering degree in Information Technology from Mumbai University. His area of research includes Data Analytics and Machine Learning.