Multi-class classification on Twitter data using LSTM neural network

Michael Shell, *Member, IEEE,* John Doe, *Fellow, OSA,* and Jane Doe, *Life Fellow, IEEE*

Abstract—The use of Social media has been used drastically, it has become a communication medium to express the emotions and opinions of an individual. Determining the sentiments regarding the product or services of the companies or different policies of the government provides the great value to the people and organization. Twitter is one of the trusted micro-blogging platform which allows its users to interact with the public by providing the options to post the videos, images with short messages. Emojis are another important feature, which has been provided by the twitter. Rather than just finding the positive or negative sentiments, multi-class sentiments also can be predicted. Therefore, in this work we are going to perform multiclass-classification for predicting multiple emotions, emojis and different sentiments from the twitter data. In this work, we have used the LSTM (Long short-term memory) model to perform multi-class classification. Where, SoftMax is used as an activation function. Data pre-processing is another important part of this work, which allows us to achieve the better results.

Index Terms—multi-class classification, twitter classification, sentiment analysis, emotion analysis, emoji analysis

1 Introduction

Twitter is one of the Most popular social media platform, where the users share their opinions, knowledge and even express their emotions for a particular event. On twitter all kind of topics, daily affair, plan, discussion and debates are shared within 280 number of characters. Due to character limitation, it becomes easier for finding out the sentiments of an individual. In order to find out the sentiments of the audience natural language processing (NLP) techniques are found to be very useful. The main purpose the applying the NLP is to understand the data from the raw text by applying mathematical and statistical operations. Sentiment analysis is one of the finest usecase of natural language processing. Finding out the sentiments and emotions from the text is highly challenging task for both human and machine.

In this research we are exploring the 3 different datasets in order to perform the sentiment analysis. These 3 dataset contains different emotions, emojis and multiple sentiments. We will perform all necessary operations to extract the useful information from it and predicting the sentiments behind it. We are gonna LSTM neural network for the implementation of the project. Accuracy and f1-score are used as performance metrics.

2 Literature Review

Various authors and researchers has performed multiple analysis on the different type of twitter data. In order to classify the sentiments from the tweets related to the electronic devices researcher has used different types of classification algorithms which includes naive bayes classifier, SVM classifier, entropy classifier and ensemble classifier. The researcher has performed feature extraction in two different steps. After their analysis they have obtained similar accuracy for every classifier for new feature vector obtained from electronic product domain tweets [1].

Another research paper publish by [2] performed sentiment analysis on tweets, where there major area of focus was analysing the customer reviews and classify their sentiments either positive, negative or neutral. Later, they have remove the neutral cases and considered this problem as binary classification problem. They have extracted the feature vectors from the twitter data and applied multiple machine learning algorithms in order to find the best performing model. They have use naive Bayes, SVM and maximum entrpy classification methods and evaluated the results in terms of precision, recall and accuracy [2]. Author Bouazizi [3] in their research stated a quantification approach, which was different from traditional multi-class classification approach. In their proposed approach they considered the tweet data which contains 11 different classes and obtained the f1-score about 45.9% [3]. In order to perform ternary classification a team of researchers has proposed SENTA tool [4] for classification to select the most fit feature from the dataset. Using their proposed approach they performed ternary classification and obtained the accuracy of 60.2%. Whereas, for binary classification the accuracy was increased to 81.3% [4] by removing the neural feature they obtained the accuracy around 71.4%.

3 Methodology

In this work we have considered 3 different datasets for our analysis, first dataset contains the tweets related to the emojis, where the data contains more than 20 different types of emojis. The second dataset contains the tweets with emotion labels, these labels can be classified as anger, joy, optimism and sadness. Our third dataset will be used to capture the sentiments of an individual by analyzing their

- *M. Shell was with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332. E-mail: see http://www.michaelshell.org/contact.html*
- *J. Doe and J. Doe are with Anonymous University.*

Manuscript received April 19, 2005; revised August 26, 2015.

tweets. We are considering the three different classes for performing sentiment analysis. The 3 different sentiment classes are positive, negative and neutral. In this work we have performed many operations on the data that can be divided into the multiple stages. The stages will be as follow.

3.1 Data Pre-processing

In the first step of data pre-processing, we are trying to extract the useful information from the text such as URL, email address, mobile numbers etc. Data cleaning is another step where we are cleaning the dataset by removing the punctuations, stop words, alphanumeric words etc for all the 3 datasets. After cleaning the dataset the cleaned files has been moved to the correct directories.

3.2 Data Visualization

In order to get familiar more with the data, we have tried to visualize the data using wordcloud and piechart. The wordcloud for emotion dataset is shown in Figure 1 As we

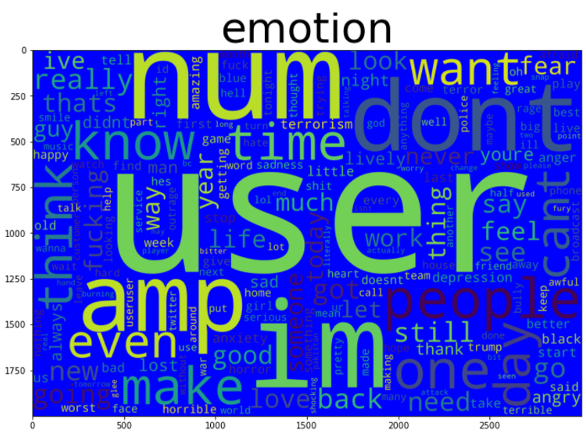

Fig. 1. Emotion Word Cloud

have collected the word dictionary from the twitter data, the word cloud has been plotted for the same, the word which will have highest number of occurrence is will have the larger font size. From the word cloud shown in Figure **??** it is clear that the occurrence of words such as user, num, im, amp, people are very high as compared to the other words. Other than the word cloud we have also analyzed the different categorize of the emotions in the data. The classification rate of emotions is shown in Figure 2.

There are multiple emotions which are represented by 0,1,2 and 3 that represents anger, Joy, optimism and sadness. It has been found in our data 42.98% tweets mainly belongs to the anger emotions. Whereas, the ration of optimism emotion is found to be very less. Further exploring the emoji dataset, we have plotted the word cloud and pie-chart in order to analyze the data in a efficient way. The word cloud obtained after analyzing the emotion dataset is shown in Figure 3.

In the Figure 3 the user word has the highest number of occurrence, followed by the word day, love and amp for emoji dataset. The different emojis can be classified into the total 20 categories, some of the highest emojis used by the twitter are red heart, smiling face, laughing etc that

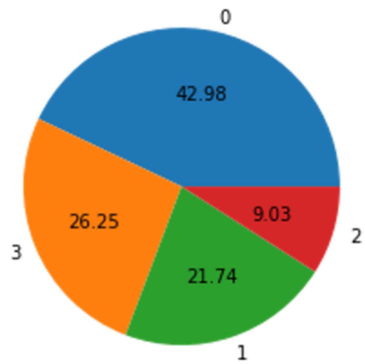

Fig. 2. Emotion Classification

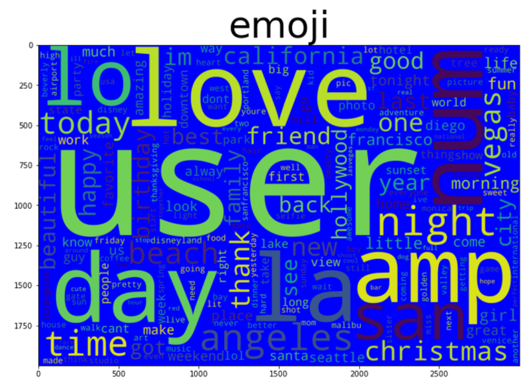

Fig. 3. Emoji Word Cloud

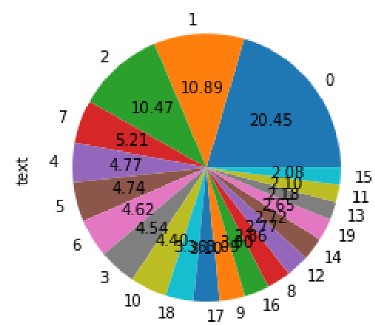

Fig. 4. Emoji Classification

basically denoted by 0, 1 and 2 in the Figure 4.

Next data exploration has been performed over sentiment dataset, the word cloud generated after analysing the sentiment twitter data is shown in Figure 5. By analyzing the word cloud for sentiment data we can say that, word 'user' has the highest number of occurrence on twitter, followed by the other words such as num, may, going, tomorrow, day etc. The sentiments are mainly classified into the 3 different classes. In the figure 6 0 represents the 'negative' sentiment, 1 represents the 'neutral' and 2 represents the 'positive' sentiment. In our dataset most of the tweets are found to be with neutral sentiments followed

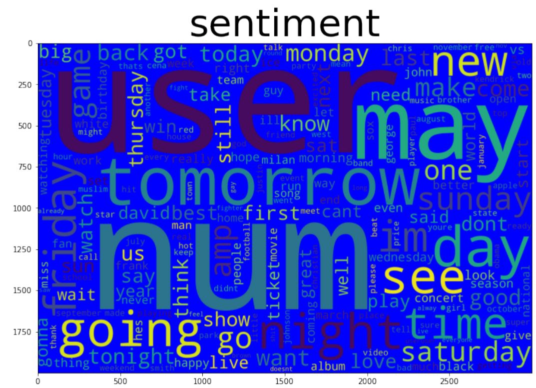

Fig. 5. Sentiments Word Cloud

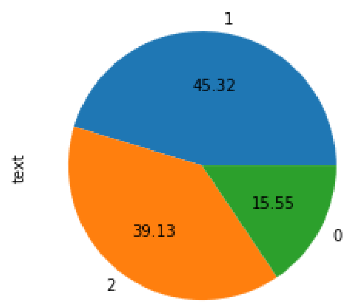

Fig. 6. Sentiment Classification

by positive sentiment and negative sentiment. The number of negative sentiments in the data are around 15.55%.

3.3 Training Data

After analysing the dataset, we have defined training, testing and validation set and used the tokenization of statements. In the process of Tokenization we are extraction the words from the sentences. After collecting the tokens we are converting every word into vector of float values by defining word2vec model. Once the vocabulary of words has been passed to word2vec model, word2vec model returns an vector of size (1X250) for every word in vocabulary. We have w2v variable, which can be considered as a dictionary and mainly contains the words as index and mean of array of every word as its corresponding values. We have also used the padding in order to ensure the dimensions of the data using pad_sequencce we are ensuring that every text line is of same length, if length of the line is less than the maximum length of the line it simply appends the pad to fit the dimensions, which will be ignored while training data. As we have seen that our target variable are categorical in nature, hence can be classified into the multiple classes. In order to handle such labels, we are using one hot encoding technique. We have used LSTM neural network algorithm for training the dataset. The LSTM(Long Short-term Memory) model mainly contains the 4 different layers the input layer, hidden layer, dense layer and output layer. We are using the softmax as an activation function.

3.4 Model Evaluation

In order to evaluate the LSTM performance we are using the accuracy and f1-score as a performance metrics. We have trained all the 3 dataset for 50 number of epochs by defining the optimizer and call back function. An early stopping function also has been used in order to avoid unnecessary model training. We have calculated both training accuracy and validation accuracy for our analysis.

4 Results

After training emotion classification data for 50 number of epoch we have achieved highest training accuracy of 42.98%. Whereas, the highest validation accuracy achieved is 39.27%. The f1-score obtained after training the model is 42.78%. The confusion matrix obtained after training the model is shown in Figure 7 .

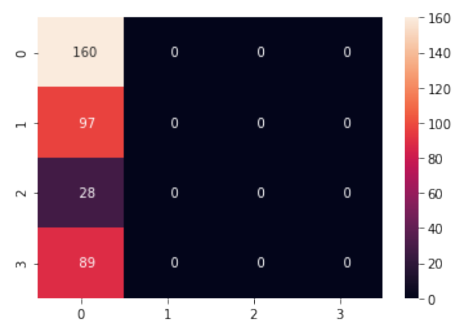

Fig. 7. Confusion matrix For Emotion Dataset

After analysing the confusion matrix shown in Figure 7.It has been found that LSTM algorithm was able to identify most of the emotions related to anger. The model was not able to find out the other classes.

In the case of emoji dataset there were more than 20 classes needs to be identified. Using LSTM neural network we have achieved highest validation accuracy of 21.60% and F1-score obtained on emoji dataset is 0.211. The confusion matrix obtained after training data is shown in Figure 8.

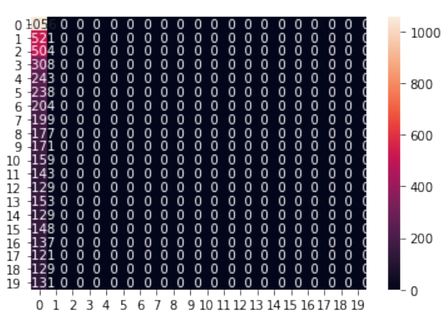

Fig. 8. Confusion matrix For Emoji Dataset

As per the analysis from the confusion matrix as shown in Figure 8 class 0, which mainly is mapped with Heart emoji was able to identified by the model. Other than that model was not able to find any other classes. After training the LSTM model for sentiment dataset we have achieved the highest validation accuracy score of 48.33% and f1-score obtained using this algorithms is 0.4345.

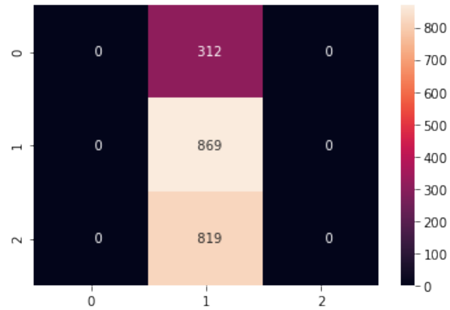

Fig. 9. Confusion matrix For Sentiment Dataset

The confusion matrix shown in Figure 9, it has been observed that class 1 which represents the neutral sentiments has been identified by the LSTM model.

5 Discussion

After analyzing the result we can say that the dataset which has the minimum number of target labels provides the better accuracy. When the number of target labels are more, a better accuracy is very difficult to achieve. In this work we have achieved lowest accuracy of 21.60% for emoji dataset. Also for the all dataset LSTM model was able to identify the one class correctly, the one of the obvious reason is the training data for the first class was high as compared to other classes in all the dataset.

6 Conclusion

Performing pre-processing on twitter data is a complex task in itself and multiclass classification adds another layers of complexity for it. Still accuracy depends on the dataset, if the dataset is balanced chances of identifying the all the classes increases. Still it depends on many factors, such as model architecture, tuning the parameters etc. Overall we have achieved the quite fair and good accuracy for all the dataset. In future work we can explore more deep learning algorithm in order to improve the results also tuning the parameteres manually can be helpful.

References

[1] M. S. Neethu and R. Rajasree, ”Sentiment analysis in twitter using machine learning techniques,” 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 2013, pp. 1-5, doi: 10.1109/ICCCNT.2013.6726818.
[2] G. Gautam and D. Yadav, ”Sentiment analysis of twitter data using machine learning approaches and semantic analysis,” 2014 Seventh International Conference on Contemporary Computing (IC3), Noida, India, 2014, pp. 437-442, doi: 10.1109/IC3.2014.6897213.
[3] Bouazizi, Mondher Ohtsuki, Tomoaki. (2018). Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer. IEEE Access. PP. 1-1. 10.1109/ACCESS.2018.2876674.
[4] M. Bouazizi and T. Ohtsuki, ”A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter,” in IEEE Access, vol. 5, pp. 20617-20639, 2017, doi: 10.1109/ACCESS.2017.2740982.