International Journal of Recent Development in Computer Technology & Software Applications
Vol. 4, Issue 2 – 2020
ISSN: 2581-6276

Stock Market Data with Deep Neural Networks using Twitter Analysis and Rumour Identification

Meda Venkata Naga Lakshmi Lavanya[1], Pokkuluri Kiran Sree[2]

[1]*PG Student, Shri Vishnu Engineering College for Women(A), Bhimavaram,*

[2]*Professor, Shri Vishnu Engineering College for Women(A), Bhimavaram,*

Abstract

In this current era, the stock market is one of the major businesses which is available online. So, for this social media plays a significant role. Increasing the percentage of people using social media for all types of information will definitely lead to the spread of rumours. In this paper, Twitter is one of the most important keywords and can be described as undoubtedly it was the most popular biggest platform in social media. Here the paper focuses on stock market data, particularly which was posted on twitter as a tweet. Here we have to identify whether that tweet can be either rumour or non-rumour. Experimenting with various machine learning and deep learning techniques BILSTM is the best method, which is the part of neural networks. By this method aggressively we can detect rumours. In this deep neural networks proposed system effectively we are using Bidirectional Long short term memory layer. In this, we are significantly discussing with four types of layers. In this BILSTM the contextual information is considered and analysis is done in two directions termed as forward and backward directions. To associate the problem of rumour identification BILSTM is the correct approach. BiLSTM already combined with LiSTM which comes under reoccurrence neural networks that work only in a unidirectional way. In this BiLSTM layer, it effectively addresses the sequences that are hidden in both forward including backwards for obtaining input.

Keywords: Deep neural networks, rumour identification, twitter analysis.

Introduction

Today the in-country stock market is one of the best developing business tools which contributes to economic growth. In this paper, we are concentrating on data classification. Deep neural networks and convolution networks can be defined in different manner. Deep neural networks are the part of an artificial network consisting more than no of layers in

© Eureka Journals 2020. All Rights Reserved.

between both layers. In this relationship can be defined in two types, neither it can be linear nor it can be nonlinear.

Neural networks are of with three layers named as input, output, hidden and each layer collectively connected with no of neurons handling weights with it. In this paper we are discussing the deep neural networks techniques like Bidirectional Long Short term memory layer gives the solution for rumour identification. Some of the problems solved by the neural networks are sales, weather forecasting, risk management up to today analysis research. Neural networks are effectively established into several types. Feedforward neural networks different from recurrent neural networks different from convolution networks. Feedforward neural networks are basic and work only in a way called unidirectional. Recurrent neural networks are complex type and with multiple directions and used for functioning improvement. convolution neural networks particularly used in image processing and face recognition. Neural networks cant destroy the system performance while doing no of tasks in a particular direction. Every neuron having the input stored inside which is not database so destroy of data won't affect here. The neuron will produce the output even any data is lost or whether that particular neuron responding or not it won't matter.

In this, we are using a Bidirectional long-short memory layer used for data classification that can be done in two ways one is forward and other is backward. In this deep neural networks, we are having four layers named as embedding layer, BiLSTM layer, DNN layer, pooling layer, flattening layer. It is level after level process done in two ways so that it can be easy for rumour detection. In this BiLSTM layer, we are having dependencies of a long term which are related to time series and input sequences.

Here we are taking the data from twitter which is having more surplus data compared to any social media platform. Mainly concentrating on the commented data in the post that is related to the stock market. So our aim is to predict that particular data is rumour or not by using some neural networks techniques.

Rumour Identification is the biggest achievement that will helps a lot of people in many ways. By lately identifying the rumour it will spread a lot of misinformation through social media and it causes a threat. Addressing the tweets very clearly and can be defined as rumour or not a rumour by using the Dense Vector formula.

Literature Survey

The research papers used in our project study are very well explained and designed with some efficiency techniques and also used to execute more algorithms. Our research papers mainly dealing with twitter data which is more accurate and using some techniques like supervised machine learning framework ,sentimental analysis, classification techniques and neural networks algorithms. Even though by using these many techniques we are missing adequate accuracy for identifying rumour in real world. Some of the drawbacks for these

© Eureka Journals 2020. All Rights Reserved.

techniques are lack of data settle and using volatility indices, text classification, multivariate regression models etc.

Including previous papers used the types of research work, we can add some features for better identification of rumours and better accuracy. Methods like more data sets, more no of attributes, more public information, statistical approach, better efficiency, good performance and any programming tool gives better accuracy for rumour identification.

Architecture

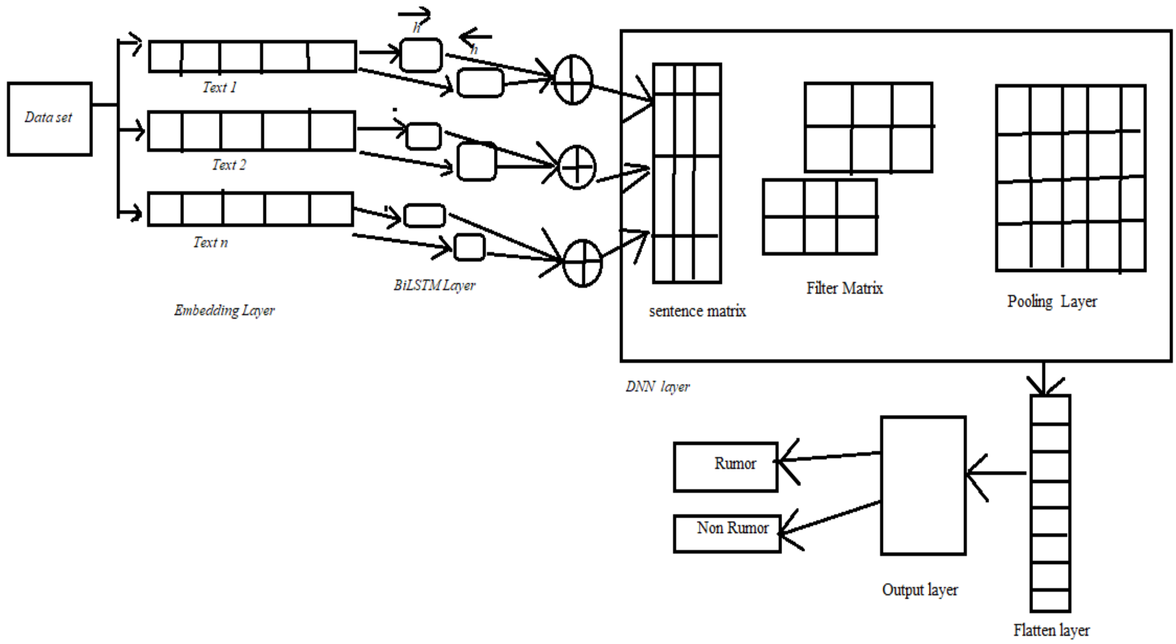

Figure 1.Proposed Architecture

Implementation

Firstly, considering one data set. From that data set, we are splitting the data into n no of parts and then gives some set of data to the next layer called BiLSTM. Here the data will comes in both directions which are used for future purposes. Evaluating those data sets together and transferred into sentence matrix. In this, each data is converted into a different separate matrix. Now should filter the matrix as per approach. Now goes to polling layer and one window size is selected and maximum matrix value is called as pooling and then flattening layer here we are using some types of functions to get clarity on rumours and then output layer finally the result will achieve in this layer.

➢ Embedding Layer

Here the data set is converting into more than of tweets. Every tweet t having some sequence of Z words denoted as x1,x2,…..xz. In this we are having some real value embedding vector called wi ∈ Rm called as word embedding. Word embedding is defined as the way of

© Eureka Journals 2020. All Rights Reserved.

representing the some types of words and set of documents as class using some dense vector representation. By determining the position the word in particular vector space called embedding using only two dimensional matrix addressing with some random weights.

- **Bidirectional LSTM Layer**

The LSTM layer can approach and obtain instructions only from earlier data in unidirectional way. To overcome that we are using BiLSTM layer in deep neural networks. Termed as Bidirectional long short term memory layer can access the information in both directions. Here BiLSTM concentrates on both past and future contexts of data. Using only sequence data particularly for classification task.

- **DNN Layer**

Defined as deep neural networks layer .In this we are having sentence matrix, filter matrix, pooling layer. In this Sentence matrix we are accessing the data that came from embedding layer and using two dimensional matrix. we are merging the data of forward direction and direction backward to create one new sentence matrix.

Filtering layer is defined as addressing some weights from their input matrix in order to generate feature map which is very useful. In this we are using element wise multiplication approach. Now combining sentence matrix with filter matrix and we have to apply some types of functions like rely activation. Adding bias value in this will help to generate appropriate value. Determining the value max (0, x) is called as rely activation

Pooling layer is the last layer in deep neural networks layer. In this we have to select some window size. Mainly pooling layer is defined as to decrease the resolution of feature map. Back propagation can also be very effective for pooling layer training. Independently pooling layer will address each and every feature map. ax pooling is the special approach in this pooling layer used to determine the maximum value in matrix.

- **Flattening Layer**

In this layer the featured matrix that came from the pooling layer is converted into separate feature vector. In flattening layer we are using sigmoid functions to generate the value with more better performance because not using any exponential operations. So here the probability is more for rumour identification. Finally calculating the overall net input value in this layer. By adding bias value will helps to more accuracy and also we are using some logarithmic value which is ‘e’ and its value is 2.46.If the net input value is greater than 0.5 it is considered as rumour and if the calculated value is less than 0.5 it is addressed as non rumour.

Algorithm

Task specific word embedding algorithmic

© Eureka Journals 2020. All Rights Reserved.

Inputs:

Dataset D,

Taking multiple no of tweets and every tweet is denoted as t,

Output:

Result shown either rumour or non rumour

Step 1: One after the other tweet should be loaded from D;

Step 2: considering the sequence of Z words,i.e.x1, x2, x3, x4…..xz.

Step 3: Real embedded vector is considered for each word of Xi where wi ∈ Rm

Step 4: Forward LSTM():

Here "x1"is the current input and sequence is x1,x2,x3…..xz-1

It will produce some sequence as output " $\vec{h}$"

Step 5: Backward LSTM():

Here "s1" is the current input and sequence is x1,x2,x3,x4….xz-1

It will produce some sequence output "$\overleftarrow{h}$"

Step 6: Now, in this step, we have to add both forward and backward sequences

output h = $\vec{h} \oplus \overleftarrow{h}$

Step 7: Now by combining those two sentence matrix is formed where H=h1,h2,h3,h4…..hz

where H ∈ Rzxm

Step 8: Now sentence matrix is aligned with a filter matrix

Step 9: By the step8 the maximum value is predicted as output from matrix output

Step 10: Finally net input is calculated, and justified as depending on the probability of the value. If probability more than 0.5 termed as rumour and otherwise termed as non rumour.

Algorithm Evaluation

Ex:-

© Eureka Journals 2020. All Rights Reserved.

International Journal of Recent Development in Computer Technology & Software Applications
Vol. 4, Issue 2 – 2020
ISSN: 2581-6276

input sentence: "*sad news plane crashed southern france*"

Embedded Layer:-

x_1	x_2	x_3	x_4	x_5	x_6
1	2	3	4	5	6

BiLSTM Layer

- **Forward LSTM**

$$\begin{array}{ccccccc} \vec{h} & & o_t & & & c_t & \\ \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix} & = & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \odot & \tau & \begin{pmatrix} 0.2 \\ 0.3 \\ 0.4 \\ 0.5 \end{pmatrix} & \end{array}$$

- **Backward LSTM**

$$\begin{array}{cccccc} \vec{h} & & o_t & & & c_t \\ \begin{bmatrix} 0.3 \\ 0.2 \\ 0.2 \\ 0.1 \end{bmatrix} & = & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} & \odot & \tau & \begin{pmatrix} 0.4 \\ 0.3 \\ 0.3 \\ 0.2 \end{pmatrix} \end{array}$$

Merged to produce a new sentence matrix H = h1, h2, h3,…..hz , H ∈ Rzxm.

$$\begin{array}{ccccc} \vec{h} & & \vec{h} & & \overleftarrow{h} \\ \begin{bmatrix} 0.5 \\ 0.5 \\ 0.6 \\ 0.6 \end{bmatrix} & = & \begin{bmatrix} 0.2 \\ 0.3 \\ 0.4 \\ 0.5 \end{bmatrix} & \oplus & \begin{pmatrix} 0.3 \\ 0.2 \\ 0.2 \\ 0.1 \end{pmatrix} \end{array}$$

DNN Layer

Matrix First Row1 = {0.5, 0.5, 0.6, 0.6}

Filtering the matrix : {0.3, 0.2, 0.2, 0.1}

Then Add to Bias = 0.57 + 0.11= 0.68

Map with Element row 1 = {0.5 * 0.3 , 0.5*0.2, 0.6*0.2 , 0.6*0.1 } = 0.57

© Eureka Journals 2020. All Rights Reserved.

Activate rely Method : max(0, 0.68)

{0.67>0} → 0.68

Flattening Layer

Maximum pool for Elements row 1 : max{0.68, 0.62, 0.46, 0.57} → selected max value is 0.68

Calculating final input:- $Y_{in\ =}$ 2.68

$Y = 1/\{1+e^{-2.68}\} = 0.9$

Y = 0.9 > ½ , Hence, current input is proved as rumour

Graph Analysis

Table 1.Task specific word embedding probability

Non-rumour	rumour
0.2	0.57
0.4	0.67
0.4	0.58

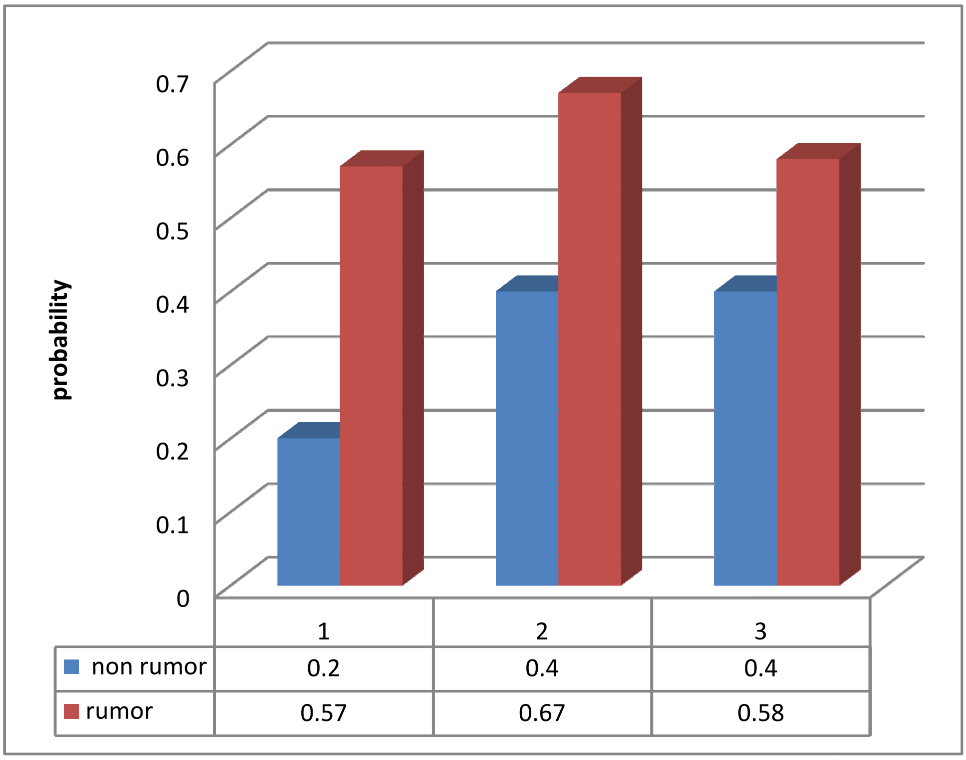

Figure 2.Task specific word embedding probability

Conclusion

© Eureka Journals 2020. All Rights Reserved.

This paper work will demonstrate the solution for rumour identification applying some neural networks model found on Bidirectional LSTM-DNN In this current analysis we are concentrating on datasets, specially twitter considering only commented facts. By this Bidirectional LSTM design. We can identify more accurately .These days rumours will spread very fast and threat is also more. So by using twitter analysis and some deep neural networks model we have identified that particular stock data is rumour or not rumour. Deep neural networks is the effective way to predict rumours and can be demonstrated very easily.

References

1. Acharya A (2017) Comparative study of machine learning algorithms for heart disease prediction.
2. Ahmad S, Asghar MZ, Alotaibi FM, Awan I (2019) Detection and classification of social media-based extremist afliations using sentiment analysis techniques. Hum Centric Compute Inf Sci 9(1):24.
3. Ahmed F, Abulaish M (2012) An mcl-based approach for spam profile detection in online social networks. In: 2012 IEEE 11th international conference on trust, security and privacy in computing and communications, pp 602-608, IEEE.
4. Alayba AM, Palade V, England M, Iqbal R (2018) A combined CNN and LSTM model for Arabic sentiment analysis. In: International cross-domain conference for machine learning and knowledge extraction, pp 179-191, Springer, Cham.
5. Asghar MZ, Rahman F, Kundi FM, Ahmad S (2019a) Development of stock market trend prediction system using multiple regression. In: Computational and mathematical organization theory, pp 1-31.
6. Asghar MZ, Ullah A, Ahmad S, Khan A (2019b) Opinion spam detection framework using hybrid classification scheme. In: Soft computing, pp 1-24.
7. Ayutthaya TSN, Pasupa K (2018) Thai sentiment analysis via bidirectional lstm-cnn model with embedding vectors and sentic features. In: 2018 International joint symposium on artificial intelligence and natural language processing (iSAI-NLP), pp 1-6.
8. IEEE Barbosa L, Feng J (2010) Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of the 23rd international conference on computational linguistics: posters, Association for Computational Linguistics, pp 36-44.
9. Ma J, Gao W, Wei Z, Lu Y, Wong KF (2015) Detect rumours using time series of social context information on microblogging websites. In: Proceedings of the 24th ACM international on conference on information and knowledge management, ACM, pp 1751-1754.
10. Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M (2016) Detecting rumours from blogs with recurrent neural networks. In: IJCAI, pp 3818-3824.

© Eureka Journals 2020. All Rights Reserved.