

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
|  | | CSPC54 AI-ML  Project Detailed Design | | | | |  | |
|  |  | | | | | | |  |
|  | | | | Group 14 |  | | | |
|  | | | | Roll Numbers106119112 106119100 Group MembersSATYARTH PANDEY RAJNEESH PANDEY SectionCSE-B |  | | | |
|  | | |  | | |  | | |

A Deep Learning Approach for Automatic Detection of Fake News

Algorithm

*Model 1 :*

**Step-1:** Pre-processing the Dataset in .*txt* format to split into Title, Content, and Label ( in Training Dataset ).

**Step-2**: Pass title and content for Word Embedding

**Step-3 :** Word Embedding :

Word embedding is a set of language modelling and feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers.

**A screenshot of a computer

Description automatically generated**

**Step 4:** After Word Embedding of inputs the resultant vectors passed to BiGRU Encoding Layer.

**Step 5 :** Bidirectional GRU’s are a type of bidirectional recurrent neural networks with only the input and forget gates.

A picture containing clock

Description automatically generatedIt allows for the use of information from both previous time steps and later time steps to make predictions about the current state.

General Structure of Bidirectional Recurrent Neural Networks

Graphical user interface, text

Description automatically generated

**Step 6 :** Attention is one **component**of a network’s architecture, and is in charge of managing and quantifying the **interdependence**:

Text

Description automatically generated

***Step 7 :***Merge Input Content and Input Title from above layers into Nodes of MLP (Multilayer Layer Perceptron):

*Text

Description automatically generated*

***Step 8 :***Five Layered MLP :

We use 512, 256, 128, 50 and 10 neurons, respectively, for five such layers with ReLU activation in each layer.

Between each such layer, we employ 20% dropout as a measurement of regularization.

Text

Description automatically generated with medium confidence

**Step 9** : Two Way Softmax Layer

Finally, the output from the last fully connected layer is fed into a final classification layer with softmax activation function having 2 neurons.

Text

Description automatically generated

**Step 10 :**

We obtain final classification Output Label as Fake or Legit

Model 2 :

**Step-1:** Pre-processing the Dataset in .*txt* format to split into Title, Content, and Label ( in Training Dataset ).

***Step 2 :***Embedding Layer: Embedding from Language Model (ELMo) :

i) Contextual i.e. representation of each word is based on entire corpus in which it is used

ii) Deep i.e. it combines all layers of a deep pre-trained neural network

iii) Character based i.e. it provides representations which are based on character, thus allowing the network to make use of morphological clues to form robust representation of out-of-vocabulary tokens during training

A screenshot of a computer

Description automatically generated

***Step 3 :***Merge Input Content and Input Title from above layers into Nodes of MLP (Multilayer Layer Perceptron):

*Text

Description automatically generated*

***Step 4*** *:* Five Layered MLP :

We use 512, 256, 128, 50 and 10 neurons, respectively, for five such layers with ReLU activation in each layer.

Between each such layer, we employ 20% dropout as a measurement of regularization.

Text

Description automatically generated with medium confidence

**Step 5 :** Two Way Softmax Layer

Finally, the output from the last fully connected layer is fed into a final classification layer with softmax activation function having 2 neurons.

Text

Description automatically generated

**Step 6 :**

We obtain final classification Output Label as Fake or Legit.

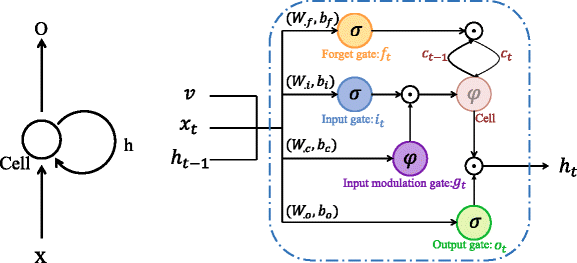
Modified Algorithm

**Step-1:** Pre-processing the Dataset in .*txt* format to split into Title, Content, and Label ( in Training Dataset ).

**Step-2 : LSTM Based architecture :**

A Bidirectional LSTMs are an extension of traditional LSTMs that can improve model performance on sequence classification problems.

Obtaining model hyperparameters using Bayesian Optimization



Text

Description automatically generated

***Note:*** *Here onwards the continuation steps are same as Previous Algorithms.*

***Step 3 :***Merge Input Content and Input Title from above layers into Nodes of MLP (Multilayer Layer Perceptron):

*Text

Description automatically generated*

***Step 4*** *:* Five Layered MLP :

We use 512, 256, 128, 50 and 10 neurons, respectively, for five such layers with ReLU activation in each layer.

Between each such layer, we employ 20% dropout as a measurement of regularization.

Text

Description automatically generated with medium confidence

**Step 5 :** Two Way Softmax Layer

Finally, the output from the last fully connected layer is fed into a final classification layer with softmax activation function having 2 neurons.

Text

Description automatically generated

**Step 6 :**

We obtain final classification Output Label as Fake or Legit

Flow Diagram of Modified Algorithm:

Diagram

Description automatically generated

Link to dataset

Official Dataset : <https://lit.eecs.umich.edu/downloads.html#Fake%20News>

Graphical user interface, text, application, email

Description automatically generated

**Note:** If the Dataset Download fails on Chrome Browser, use Firefox Browser or below link.

Above Dataset uploaded on G-Drive : <https://drive.google.com/drive/folders/1v1LDcvGZhBV-Ffq1uPfeyAT2NVrtORFf?usp=sharing>

Tools to be used.

**Programming Language:** Python

**Frameworks :** TensorFlow, Keras.

**Implementation Environment:** Jupyter Notebook