e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Conference:01/December-2022 Impact Factor- 6.752 www.irjmets.com
1st National Conference on Applications of soft Computing Techniques in Engineering NCASCTE-2022
Organized by Department of Electrical & Electronics Engineering, ACE Engineering College, Hyderabad

NEXT WORD PREDICTION

D. Ashwini*1, M. Karthik*2, V. Sampath*3, M. Nandini*4

*1Associate Processor, Department Of Computer Science Engineering, ACE Engineering College, Ghatkesar, Telangana, India.

*2,3,4Student, Department Of Computer Science Engineering, ACE Engineering College, Ghatkesar, Telangana, India.

DOI: https://www.doi.org/10.56726/IRJMETS-NCASCTE202222

ABSTRACT

A program, the aim of which is to give suggestions of possible next words guessed on the basis of the user's input. When a user is typing text on a mobile device it can be useful to suggest the next word as this will optimize typing time and also avoid possible errors. However, this data has private information, which limits its movement to a centralized environment. It is one of the fundamental tasks of NLP and has many applications. This article deals with how we can use a neural model better than a basic RNN and use it to predict the next word. We deal with a model called LongShort-term Memory (LSTM). We can use the TensorFlow library in python for building and training the deep learning model.

Keywords: NLP, RNN, LSTM & TensorFlow.

I. INTRODUCTION

Natural Language Processing (NLP) is a significant part of artificial Intelligence, which incorporates AI, which contributes to finding productive approaches to speak with people and gain from the associations with them. One such commitment is to give portable clients anticipated" next words," as they type along within applications, with an end goal to assist message conveyance by having the client select a proposed word as opposed to composing it. As LSTM is Long short time memory it will understand the past text and predict the words which may be helpful for the user to frame sentences and this technique uses a letter-to-letter prediction means it predicts a character to create a word. As writing an essay and framing a big paragraph are time-consuming it will help end-users to frame important parts of the paragraph and help users to focus on the topic instead of wasting time on what to type next. We expect to create or mimic auto-complete features using LSTM. Most of the software uses different methods like NLP and normal neural networks to do this task we will be experimenting with this problem using LSTM by using the Default Nietzsche text file also known as our training data to train a model. Next Word Prediction is also called Language Modelling that is the task of predicting what word comes next. It is one of the fundamental tasks of NLP and has many applications.

II. METHODOLOGY

Natural language processing has been an area of research and used widely in different applications. We often love texting each other and find that whenever we try to type a text a suggestion poops up trying to predict the next. word we want to write. This process of prediction is one of the applications NLP deals with. We have made huge progress here and we can use Recurrent neural networks for such a process. There have been difficulties in basic RNN. This article deals with how we can use a neural model better than a basic RNN and use it to predict the next word. We deal with a model called Long Short-term Memory (LSTM). We can use the TensorFlow library in python for building and training the deep learning model. Vanishing gradient descend is a problem faced by neural networks when we go for backpropagation as discussed here. It has a huge effect and the weight update process is widely affected and the model became useless. So, we used LSTM which has a hidden state and a memory cell with three gates that are forgotten, read, and input gate.

EXISTING SYSTEMS

Google also uses the next word prediction model based on our browsing history. Google uses our browsing history to make next-word predictions, smartphones, and all the keyboards that are trained to predict the next word are trained using some data.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Conference:01/December-2022 Impact Factor- 6.752 www.irjmets.com
1st National Conference on Applications of soft Computing Techniques in Engineering NCASCTE-2022
Organized by Department of Electrical & Electronics Engineering, ACE Engineering College, Hyderabad

Microsoft Outlook is a great email and calendar program, but it doesn't have the right predictive text to make it easier to type text.

The program only offers four different predictive texts that can help you type properly.

Once Text Predictions is enabled, you can start typing and the system will suggest possible matches. You can just tap on any suggestion to send your message.

The text prediction feature, which is turned on by default, anticipates your next words while writing/typing and suggests words or phrases as you type. You can accept or reject the suggested word or phrase.

To accept a suggested word/phrase, press the Tab key or right-arrow key, and to reject the same, you can press the Esc key or keep typing the word/phrase you have in your mind.

The feature indeed makes composing documents easier and faster. However, not all users like to see word/phrase suggestions while writing.

Figure 1: Google using next word prediction

III. MODEL AND ANALYSIS

Next-word prediction is the most widely used machine-learning technique in real-life applications. There are efficient models which give higher accuracy while predicting the output. Our model predicts the next word and can be used to form a new sentence and an entire paragraph that is meaning full to read. An input of 50 words is given to the machine learning model to predict the next word. The predicted word is added to the next input and continues this process until the required paragraph is produced. Every time a new input is given and the output of the previous input is added the first word of the input is removed which makes our model different from existing models so which decreases the use of words to predict the next word to increase efficiency.

We have used TensorFlow, Keras and LSTM which is a natural language Processor.

TensorFlow-TensorFlow allows developers to create dataflow graphs—structures that describe how data moves through a graph, or a series of processing nodes. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array, or tensor.

TensorFlow applications can be run on most any target that's convenient: a local machine, a cluster in the cloud, iOS and Android devices, CPUs or GPUs. If you use Google's own cloud, you can run TensorFlow on Google's custom TensorFlow Processing Unit (TPU) silicon for further acceleration. The resulting models created by TensorFlow, though, can be deployed on most any device where they will be used to serve predictions.

LSTM- Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was designed by Hochreiter&Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Conference:01/December-2022 Impact Factor- 6.752 www.irjmets.com
1st National Conference on Applications of soft Computing Techniques in Engineering NCASCTE-2022
Organized by Department of Electrical & Electronics Engineering, ACE Engineering College, Hyderabad

not give an efficient performance. LSTM can by default retain the information for a long period of time. It is used for processing, predicting, and classifying on the basis of time-series data.

IV. RESULTS & DISCUSSION

As mentioned in the model and analysis the data set we used is William Shakespeare's sonnet from the 100th etext file presented by Project Gutenberg. There are over lakhs of words present in that file. The dataset is inserted into the variable divided by "\n" for each new line. The output looks like "This is the 100th Etext file presented by Project Gutenberg, and\nis presented in cooperation with World Library, Inc., from their\n" .The resulting output where "\n" is present is joined by a comma and forms single paragraphs of lakhs of words. Later the entire dataset is filtered by removing punctuations and duplicate words for removing redundancy.

The model architecture is shown in the figure below. To plot the model, here used to. Keras.utils.plot_model. Then to show the layer name, just need to change the value of show_layer_names to True. To download the model and name it, need to use to_file and then just name the file. Here researchers used this simple model so it won't take a long time when being run.

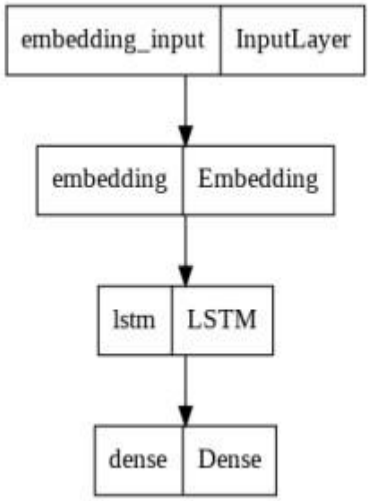

Figure 2. Architectural model.

To show the model summary, just need to call the model. The summary then the result will be shown. Here's the model summary of the model which here it's using a sequential model.

```
model.summary()

Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_1 (Embedding)     (None, 50, 50)            122450

 lstm_2 (LSTM)               (None, 50, 100)           60400

 lstm_3 (LSTM)               (None, 100)               80400

 dense_2 (Dense)             (None, 100)               10100

 dense_3 (Dense)             (None, 2449)              247349

=================================================================
Total params: 520,699
Trainable params: 520,699
Non-trainable params: 0
```

Figure 3. Model Summary.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Conference:01/December-2022 Impact Factor- 6.752 www.irjmets.com
1st National Conference on Applications of soft Computing Techniques in Engineering NCASCTE-2022
Organized by Department of Electrical & Electronics Engineering, ACE Engineering College, Hyderabad

To train the model, here use model fit with 3 parameters: its xs,ys, and the epochs so that it could show the full history of the training. The epochs are 200. The last result in epoch 200/200 is 8ms/step, loss: 55%, and accuracy: 75%. Compared to the two other reaseaher that has mentioned in the introduction, this accuracy is good enough as it managed to get 75%. While the other research, say the one using RNN just got around 54%-55% and the one using the Pre Training Federated Text Model got around 21%-22%. The difference of the result could happen because there are some differences such as in the dataset, code and for sure the model.

E. Graph

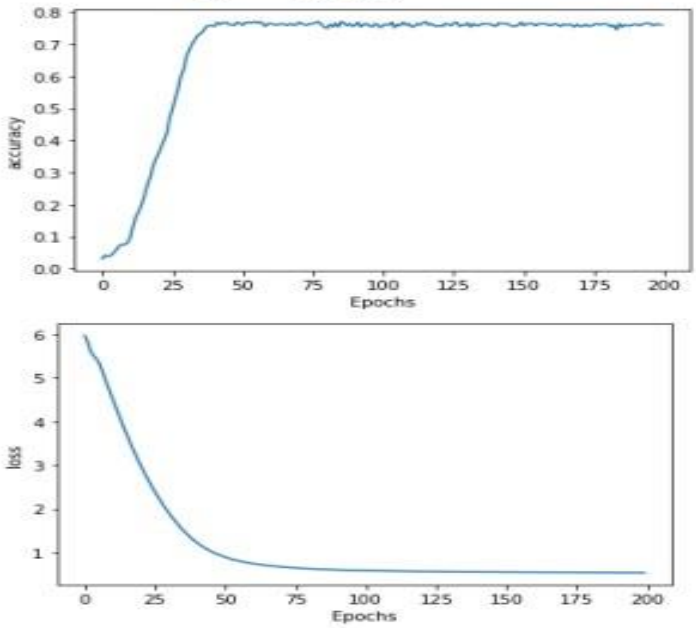

Figure 4. Accuracy and Loss graph

To make predictions as the test to show the implementation of the built model, here we need to make a function which can predict the next word. The parameters are the input of what destination is looking for by the user and how many next words want to be predicted. Here's an example of what the input and output looks like.

```
seed_text="from fairest creatures we desire increase that thereby beautys rose might never die but as the riper shou

generate_text(model,tokenizer,seq_length,seed_text,200)

'and found example to thee are be my self shall be forgot thou art but thou art slain i knowst thee my soul of my he
t flies and heavens air doth some respects are pale be the prey of thee the very main of time join for the world anc
akness doth see giving if wantonness lived with be men but like for thee are love the self is i be bright'
```

Figure 5. input and output predicted.

V. CONCLUSION

Consider, you have the task of modifying certain information in a calendar. To do this, an RNN completely changes the existing data by applying a function. Whereas, LSTM makes small modifications on the data by simple addition or multiplication that flow through cell states. This is how LSTM forgets and remembers things selectively, which makes it an improvement over RNNs. Now consider, you want to process data with periodic patterns in it, such as predicting the sales of colored powder that peaks at the time of Holi in India. A good strategy is to look back at the sales records of the previous year. So, you need to know what data needs to be forgotten and what needs to be stored for later reference. Else, you need to have a really good memory. Recurrent neural network seems to be doing a good job at this, theoretically. However, they have two downsides, exploding gradient and vanishing gradient, that make them redundant. Here, LSTM introduces memory.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Conference:01/December-2022 Impact Factor- 6.752 www.irjmets.com
1st National Conference on Applications of soft Computing Techniques in Engineering NCASCTE-2022
Organized by Department of Electrical & Electronics Engineering, ACE Engineering College, Hyderabad

ACKNOWLEDGEMENT

We would like to thank our guide **Mrs.D.Ashwini**for his continuous support and guidance. Also, we are thankful to our project coordinator **Mr. K. Prem Kumar** &**Mr. K. Lal Bhahudhur**and we are extremely grateful to **Dr. M. V. Vijaya Saradhi**, Head of the Department of Computer Science and Engineering, ACE Engineering College for his support and invaluable time.

VI. REFERENCES

[1] Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." Science 349, no. 6245 (2015): 255-260.

[2] Sahoo, Abhaya Kumar, Chittaranjan Pradhan, and Himansu Das. "Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making." In Nature inspired computing for data science, pp. 201- 212. Springer, Cham, 2020.

[3] Prajapati, Gend Lal, and Rekha Saha. "REEDS: Relevance and enhanced entropy based Dempster Shafer approach for next word prediction using language model." Journal of Computational Science 35 (2019): 1-11.

[4] Ambulgekar, Sourabh, Sanket Malewadikar, Raju Garande, and Bharti Joshi. "Next Words Prediction Using Recurrent NeuralNetworks." In ITM Web of Conferences, vol. 40, p. 03034. EDP Sciences, 2021.

[5] Stremmel, Joel, and Arjun Singh. "Pretraining federated text models for next word prediction." In Future of Information and Communication Conference, pp. 477-488. Springer, Cham, 2021.

[6] Z. Shi, et. al.have defined that recurrent neural network has input, output and hidden layer. The current hidden layer is calculated by current input layer and previous hidden layer. LSTM is a special Recurrent Neural Network. The repeating module of ordinary RNN has a simple structure instead LSTM uses more complex function to replace it for more accurate result. The key element in the LSTM is the cell state which is also called as hidden layer state.

[7] J. Shin, et. al. have defined that understanding the contextual aspects of a sentence is very important while its classification .This paper mainly focuses on it. Various approaches like SVM, T-LSTM, and CNN have been previously used for sentence classification. But, the proposed C-CNN (Contextual-CNN) gives better results i.e. The C-CNN achieves state-of-theart accuracy 52.3% on the fine-grained sentiment prediction task and 95.2% on the TREC question categorization task.