IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11773

NEXT WORD PREDECTION USING MACHNIE LEARNING

VISHWAS D. K[1], Prof K.M SOWMYA SHREE[2]

PG Scholar, Dept. of MCA, PES College of Engineering, Mandya, Karnataka, India[1]

Assistant Professor, Dept. of MCA, PES College of Engineering, Mandya, Karnataka, India[2]

Abstract: Next word prediction using machine learning which includes 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.

Keywords: Predection, Information

1. INTRODUCTION

With the big data revolution, old problems demanded new ways to solve them, and most especially, we find Deep Learning (DL). One of the domains affected by this revolution is Natural Language Processing (NLP), where with the rapid development and proliferation of social media and websites we have a massive amount of data that can help us to achieve good results in this field. One of NLP sub-domains is Language Modeling (LM) which includes the task of word prediction.

With the technological development and the great use of electronic devices, the search for ways to speed up the process of entering information became an imperative, as the beginnings were since the emergence of augmentative and alternative communication systems, as well as word prediction systems since 1980. To face the next word prediction problem, the beginnings were with using statistical and probabilistic methods, then combined with knowledge-based models, then heuristic modeling. After the immediate surge in data, machine learning techniques stepped in to suggest good solutions and finally, with the strong return of neural networks because of its revolutionary results with big data.

2. LITERATURE SURVEY

S. Lai, have proposed the setting based data order; RCNN is extremely helpful. The presentation is best in a few datasets especially on report level datasets. Contingent upon the words utilized in the sentences, loads are doled out to it and are pooled into least, normal and the maximum pools. Here, max pooling is applied to extricate the catchphrases from the sentences which are generally significant. RNN, CNN and RCNN when contrasted and other conventional techniques, for example, LDA, Tree Kernel and calculated relapse creates high exact outcomes. A. Hassa, have proposed RNN for the construction sentence portrayal. This tree like design catches the semantic of the sentences. The text is broke down word by word by utilizing RNN then the semantic of the relative multitude of past texts are safeguarded in a proper size stowed away layer. For the proposed framework LSTM assumes significant part, being a memory stockpiling, it holds the characters which helps in anticipating the following word. [1].

Tran have suggested that n-gram is a touching succession of 'n' things from a given grouping of text. Assuming that the given sentence is 'S', we can develop a rundown on n-grams from 'S', by finding sets of words that happens close to one another. The model is utilized to infer likelihood of sentences utilizing the chain rule of unrestricted likelihood. Z. Shi have characterized that intermittent brain network has info, yield and secret layer. The ongoing secret layer is determined by current information layer and past secret layer.[2]

LSTM is an extraordinary Recurrent Neural Network. The rehashing module of common RNN has a straightforward design rather LSTM utilizes more perplexing capability to substitute it for more precise outcome. The vital component

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11773

in the LSTM is the phone state which is additionally called as covered up layer state.J. Shin, et. al. [10] have defined that understanding the contextual aspects of a sentence is very important while its classification. This paper mainly focuses on it. Different methodologies like SVM, T-LSTM, and CNN have been recently utilized for sentence arrangement.[3]

Sutskever, have defined deep learning being the newest technology in the era has advanced in many fields. One of the techniques called as Deep Neural Networks are very powerful machine learning models and have achieved efficient and excellent performance on many problems like speech recognition, visual object detection etc. due to its ability to perform parallel computation for the modest no of steps. Many attempts have been made to address the problems with neural network. The results showed that a large deep LSTM with a limited vocabulary can outperform a standard SMT-based system.[4]

3. EXISTING SYSTEM

This course of foreseeing the following Word is very perplexing in light of the fact that we need to anticipate words which client suspect as much it is anticipating the contemplations of the client so the exactness is very low contrast with another ML projectCalculation similar to SVM, Decision Tree, etc will not provide better results and it will take more time to predict the outputs as the process is quite complex.

To wake useful predictions, b text predictor needs a knowledge about 1anguage as possible so we should keep training the neural network continue on the new 1anguages and the new data.

4. PROPOSED SYSTEM

This proposed work is an outline to create an adoptable model that can assist clients with recognizing next word while figuring out client vocable in a quick and powerful way so client need to give 40 letters then it passes this 1etter to LSTM NN and predicts N number of 1etters.

LSTM get it and gain proficiency with every single letter, letter by letter and make a score to the following letter.

Which This score on the other hand it will go through a similar LSTM and later it will foresee a word letter by letter.Which it contain the different modules were it extract the words from the image and it will preprocess the prediction and identify the next word

5. SYSTEM ARCHITECTURE

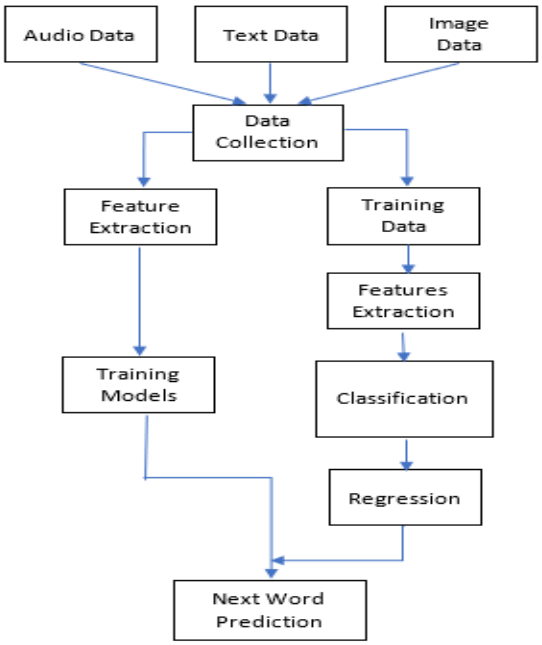

Fig 1 Architecture of Predection

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022
DOI: 10.17148/IJARCCE.2022.11773

6. CONCLUSION

Thus, in this project we have constructed a message indicator utilizing LSTM, you can likewise deal with other NLP applications like opinion analyser or message classifier. The exactness of the model can be improved by tuning the hyperparameters, expanding implanting layers or rolling out certain improvements in the information pre-handling.
Word conjecture is an imperative end eavor and has various basic app1ications. A strong word figure system can help clients, by allowing higher text entry rates, and restricting number of typ0graphical mix-ups and inaccurate spellings. This point has been seen by the architects of the open-source word processor 0penOffice, which gives, close by standard word dealing with features, word finish (www.openoffice.org). Later on headings of this investigation, we should endeavor several new perspectives to additionally foster the figure accuracy furthermore. For examp1e, we will look at extending the setting size without impacting the computation complexity of the methodology. Similarly, we plan to research the possibility including positiona1 information ab0ut the setting features in the developing experience.
.

FUTURE ENHANCEMENT

In the proposed system there is scope for improvement. Following are some of the enhancement.

- we propose utilizing a huge data set or a few information bases to cover the biggest conceivable number of vocabularies, as well as changing boundary values or utilizing a bigger organization.
- We likewise propose to work with explicit space (sport, news,…) datasets, this can diminish the jargon size and gives more engaged results. The investigation of some new profound learning stages for cell phones can likewise be proposed as an augmentation.

REFERENCES

[1] Bhuyan M P, Sarma S K. (2019).An N-gram based model for predicting of word-formation in Assamese language [J]. Journal of Information & Optimization Sciences, 40(2):427-440.
[2] Hamarashid H K, Saeed S A, Rashid T A, et al. (2021). Next word prediction based on the N-gram model for Kurdish Sorani and Kurmanji [J]. Neural Computing and Applications, 33(5): 4547-4566.
[3] Le C C, Prasad P, Alsadoon A, et al. (2019). Text Classification: Naieve Bayes Classifier with Sentiment Lexicon[J]. IAENG Internaitonal journal of computer science, 46(2):141-148.
[4] Atlinar F, Ayar T, Darrige A, et al. (2020). Masked Word Prediction with Statistical and Neural Language Models[C]// 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 1-4
[5] Nicole M, Klinton B. (2019). Using LSTMs to Assess the Obligatoriness of Phonological Distinctive Features for Phonotactic Learning[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 1595-1605.

TEXT BOOKS

1. Java: A Beginner's Guide, Sixth Edition by Herbert Schildt.
2. Head First Android Development: ABrain-Friendly Guide 1st Edition by Dawn Griffiths
3. Android Programming for Beginners: By John Horton
4. Android App Development for Dummies by Michel Burton/