

TEAM 21

NEXT WORD PREDICTION

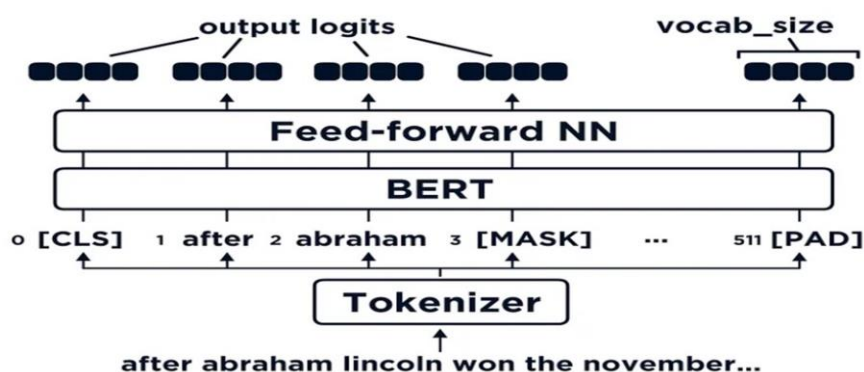
Abstract

The system displays a list of the most probable words that could appear in the given position. The project is implemented using three language models, i.e., LSTM (Long Short-Term Memory), BERT (Bidirectional Encoder Representations from Transformers) and GRU (Gated Recurrent Unit) to experiment and check statistics for expecting subsequent words. SWAG (Situations With Adversarial Generations) dataset has been used to train the models and acquire the output. The model uses large scale training data to learn contextual embeddings of words and sentences. The main aim of our project is to predict next relevant word using BERT model.

Modules

- Data Collection and Preprocessing
- Train and Test Split
- Model Training
- Model Selection
- Integrating GUI

Architecture



Tools and Technologies

- Tensorflow and keras
- NLTK
- Transformers
- Streamlit
- Numpy
- Pandas
- Matplotlib

Conclusion and Future Scope

As a result of BERT, LSTM, and GRU models, it is determined that the BERT model gives more successful results on SWAG corpus. With BERT model, higher accuracy and lower loss is observed when compared with the other models. With more equipped and powerful computation based system can provide more accurate results when trained on large corpus.

Guide Name

Dr. Aruna Rao S L
 Professor & HOD
 it.hod@bvrithyderabad.edu.in

Team Members



K.Sai Tejaswi
20WH1A1206



M.Sandhya
20WH1A1243



Ch.Sriniketha
20WH1A1255

Github links

1. <https://github.com/Saitejaswi-K/Next-Word-Prediction>
2. [https://github.com/Sandhya0518/20WH1A1243-Next Word Prediction](https://github.com/Sandhya0518/20WH1A1243-Next_Word_Prediction)
3. <https://github.com/SrinikethaCh/Next-Word-Prediction>