BVRIT HYDERABAD
College of Engineering for Women
Department of Information Technology
Major Project - Academic Year 2019-20

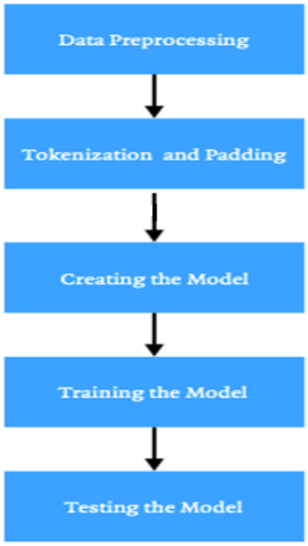
TEAM 02

NEURAL MACHINE TRANSLATION FOR INDIAN LANGUAGES

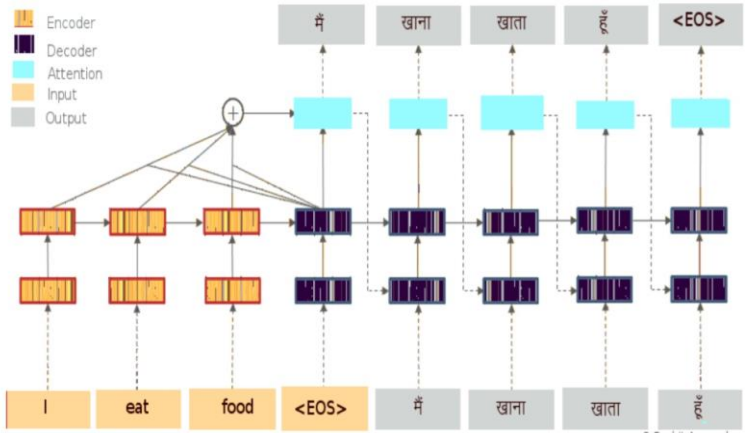
Abstract

Neural Machine Translation (NMT) is a variation of machine translation system. It learns how to translate data itself by using training data. This translation is based on the two ideas, one is Recurrent Neural Network, and another is Encoding. Recurrent Neural Network takes the inputs, along with this input it will take previous output as a input. Finally gives the output by taking all these inputs. This output is an input for next translation. This project aim is to create a Neural Network to translate from English to Hindi. For this an extension of RNN (Recurrent Neural Network) i.e. LSTM (Long Short-Term Memory) is being used then. This project is implemented in six modules they are as follows: Data Preprocessing, Tokenization and Padding, Word Embeddings, creating a model, training a model and Testing a model. This LSTM was able to translate English input sentences to Hindi output sentences.

Modules



Architecture



Tools

- Python 3
- Keras with backend as Tensorflow
- Jupyter Notebook

Conclusion & Future Scope

India is a linguistically rich and diverse country, so there will always be a requirement for translation from one language to another language. Machine Translation, translates text from the English language to Hindi language text in the domain of administration which translates documents such as government appointments, circulars, office orders, etc. There are mainly three approaches for Machine Translation. They are Linguistic or rule-based, Non-Linguistic and Hybrid. However, these methods have limitations in terms of scope and efficiency. That is, these can translate when input sentences are grammatically and syntactically correct, and the translated output may not be in commonly used language/style. The future scope of the project is to improve the performance and accuracy of the output sentence while working with a large dataset.

Guide

Mr. L. Naveen Kumar
Assistant Professor of I.T.
Email: naveenkumar.l@bvrithyderabad.edu.in

TEAM




A Bavishya 16WH1A1202
G DivyaLata 16WH1A1215
V BhanuSamera 16WH1A1247
B Sandhya 16WH1a1210