Data Mining and Predictive Analysis Lab
Project results
Semester 6 - CCE B

Machine Translation using Transformers

I. **Introduction:**

This project is an implementation of the paper - “Attention is all you need” (https://arxiv.org/abs/1706.03762). We sought to apply the transformer model architecture introduced in this paper to translate sentences from Hindi to English and vice-versa. Currently, we have achieved basic performance for this model and can translate most of the commonly used phrases and sentences. We chose this paper since it proved to be a breakthrough in the field of NLP and wanted to extend its functionality to machine translation.

II. **Dataset insight:**

- No. of Hindi - English pairs: 6,90,722
- Sources: IITB Text Corpus and Indic Languages Parallel Corpus
- Maximum sentence length: 64 words

III. **Training statistics:**

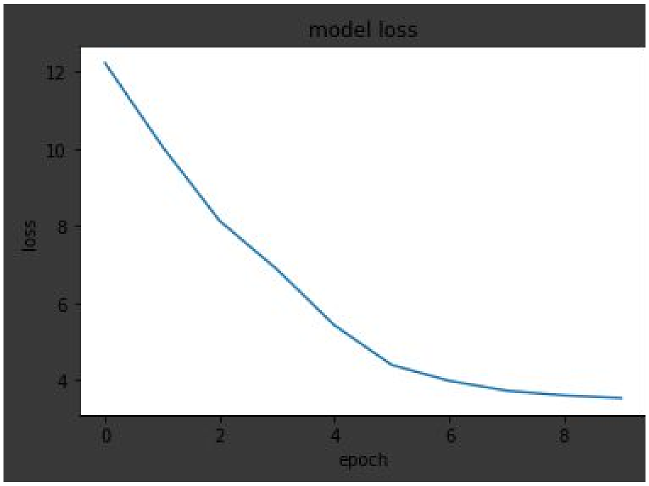
Hyperparameters - Loss algorithm: Categorical Crossentropy, Optimizer: Adam, Batch size: 100, Learning rate: 0.003, Total model parameters: 52,614,811, Prediction algorithm: Greedy based approach, Training time per epoch ~ 2 hours.

1. **English - Hindi:**
   - No. of epochs: 8
   - Total training time ~ 16 hours
   - Training Loss: 3.53
   - Validation Loss: 3.27

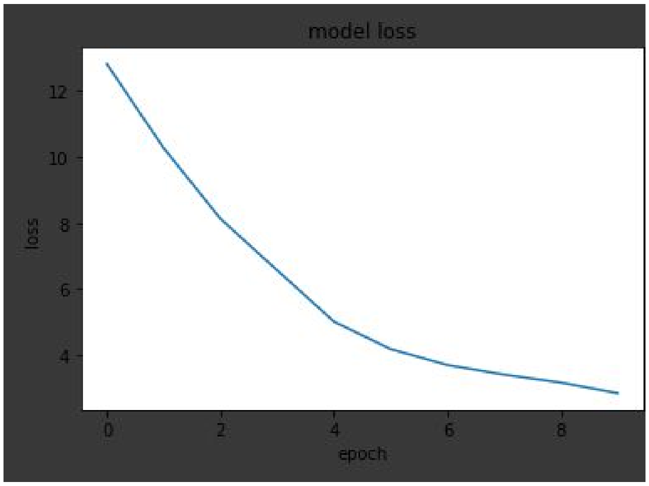
2. **Hindi - English:**
   - No. of epochs: 10
   - Total training time ~ 20 hours
   - Training Loss: 2.86
   - Validation Loss: 2.93

IV. Visualization

The steady decrease in the model loss for both cases is shown below:



Hindi - English



English - Hindi

V. Model Predictions

1. English - Hindi

a. Few sentences were chosen at random from the training dataset and the performance of the model was recorded. Here are the results:

Input	Expected	Prediction
The real privilege of leadership.	नेतृत्व का सच्चा विशेषाधिकार	नेतृत्व का वास्तविक सम्मान
Within three years, the city emerged as the main citizen center.	अगले तीस वर्षों में नगर एक प्रधान नागरिक केंद्र के रुप में विकसित हुआ ।	तीन वर्ष के दौरान नगर राज्य के मुख्य केंद्र में आया ।
Click to close the side pane.	बाजू फलक को बंद करने के लिए क्लिक करें	बाजू फलक को बंद करने के लिए क्लिक करें
Its fertile land yielded two crops a year.	इसकी उपजाऊ भूमि में साल में दो फसलें पैदा होती थीं ।	दो वर्ष तक यह फसल फसल का उत्पादन हुआ ।
An area within which one can only operate .	ऐसा क्षेत्र जिसमें केवल वही एक कार्य करने को सक्षम हो ।	एक क्षेत्र जो केवल एक क्षेत्र के भीतर ही काम कर सकता है ।

b. An additional 5 sentences were generated outside the training dataset to test the performance of the model. Here are the results:

Input	**Prediction**
Can you buy some flowers?	क्या आप कुछ फूल खरीद सकते हैं ?
I have to go back to my home now.	अब मैं घर जाने जा रहा हूँ
The Indian army succeeded in defeating their enemy.	भारतीय सेना ने अपने आक्रमण में सफलता प्राप्त की ।
Can you assist me in the temple chores?	क्या आप मंदिर में सहायता कर सकते हैं ?
There are many grape trees	कई अंगूर वृक्ष हैं ।

2. **Hindi - English**

a. Few sentences were chosen at random from the training dataset and the performance of the model was recorded. Here are the results:

Input	**Expected**	**Prediction**
वे अपने हाथों का इस्तेमाल कर रहे हैं ।	They 're using their hands	They are using their hands.
किसी वस्तु का केंद्रीय भाग	Central part of an object	The central part of a commodity.
फ़ाइल प्रबंधक में विशिष्ट माध्यम के लिए फ़ोल्डर खोलता है	Opens the folder for a specific medium in the file manager	Open folder for the specific folders in the file manager.
एकल बर्स्ट में लेने के लिए तस्वीर की संख्या	The number of photos to take in a single burst.	The number of photos to show in single burst.
पैराग्राफ की एक सूची प्रदर्शित करता है	Displays a list of paragraphs.	Displays a list of paragraph.

b. An additional 5 sentences were generated outside the training dataset to test the performance of the model. Here are the results:

Input	**Prediction**
तुम मेरी बात सुनते क्यों नहीं हो?	Why don 't you hear me?
हाथी अफ़्रीका में रहते हैं।	Elephants are living in africa
मैं एक सफल व्यापारी हूँ	I am a successful employee
क्या मुझे आपके लिए खाना बनाना चाहिए?	Should i make you food?
क्या मैं आपकी मदद कर सकता हुँ ?	Can i help you?

VI. Project Code

The entire project has been developed on a Google Colab Notebook. The prerequisite files and the code has been uploaded to Github. The link for the same is: https://github.com/Chinnu1103/Hindi-English-Translator-using-Transformers

VII. Team Member Details

Roll No.	**Name**	**Reg. No.**
20	Chirangivi Bhat	170953082
30	Aakash Vashishtha	170953124
34	V. Advaith	170953148