Enhancing Machine Translation using Multi-head Self Attention with Bidirectional LSTM & fine-tune Hugging face T5 Model,Alignment head

Team Members:

Bahareh Arghavani
Naveen Preetham Kaveti
Nagabhushan Reddy Kadiri

Advisor:
Khaled Sayed

Introduction

We aim to improve the performance of our current machine translation system. The presentation focuses on the following models:

- Our proposed model which is a multi-head self attention with a bi-directional LSTM in translation part (encoder) and LSTM in language model (decoder)

- The open-source Hugging face T5 model and Transformer Align Model languages

Both models will be applied to English to French, English to German and for English to Hindi, and Farsi to English we used Transformer Align Model languages.

Project Objectives

Develop a robust neural machine translation model for sequence-to-sequence tasks, addressing challenges in capturing long-range dependencies and improving translation accuracy.

Approach

- **Integration of bidirectional LSTMs and multi-head self-attention.**
- **Utilization of dropout regularization for overfitting prevention.**
- **Sequential refinement through multiple Bidirectional LSTM layers in encoder and single LSTM layer in decoder .**
- **Output alignment using RepeatVector.**

Architecture Overview

Input Stage:

Input sequences embedded via an embedding layer.

Bidirectional LSTMs process embedded sequences.

Attention Mechanism:

Multi-head self-attention captures long-range dependencies.

Dropout layer introduced for regularization.

Architecture Overview

Sequence Refinement:

Second bidirectional LSTM refines attended sequences.

RepeatVector aligns output with desired sequence length.

Final unidirectional LSTM captures output dependencies.

Output Stage:

Dense output layer with softmax activation for translation probabilities.

Model compiled with RMSprop optimizer and sparse categorical cross-entropy loss.

Architecture Overview

Key Features:

Bidirectional processing for comprehensive context.

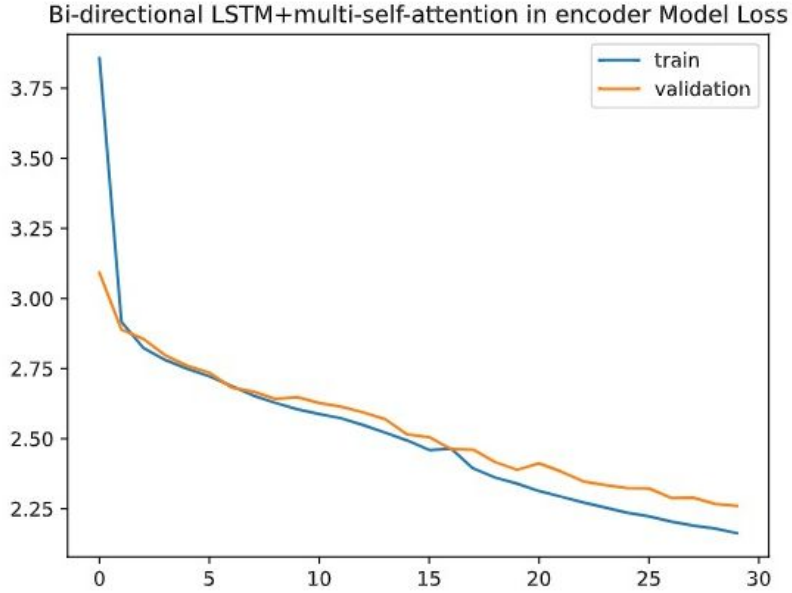
Multi-head attention for nuanced dependency capture.

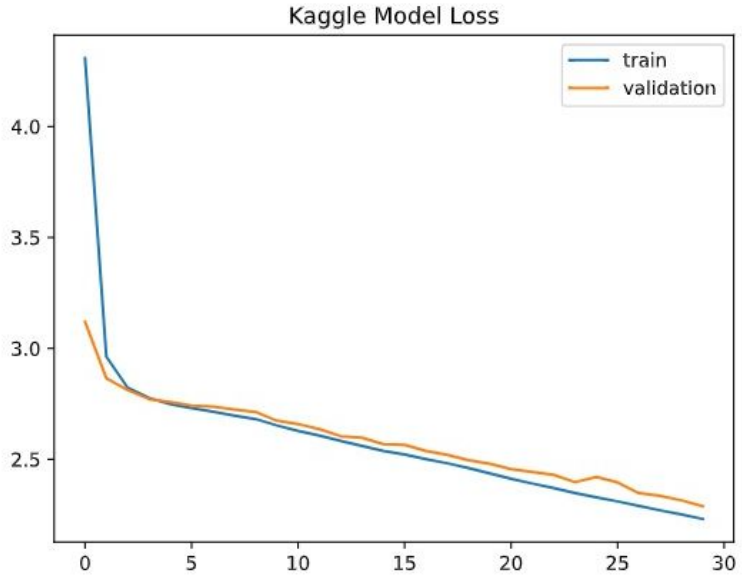
Dropout and Repeat Vector for regularization and output alignment.

Optimized compilation for effective training.

This model architecture is poised to enhance the state-of-the-art in neural machine translation, combining the strengths of bidirectional LSTMs and multi-head self-attention to achieve superior accuracy and flexibility in handling diverse sequence lengths.

Architecture Overview





Hugging Face T5 model

T5 (Text-To-Text Transfer Transformer) reframes all NLP tasks into a unified text-to-text-format. It’s a robust language model with 60 million parameters and can handle different NLP tasks using the same model structure, loss function, and hyperparameters.

Implementation of the T5 Model

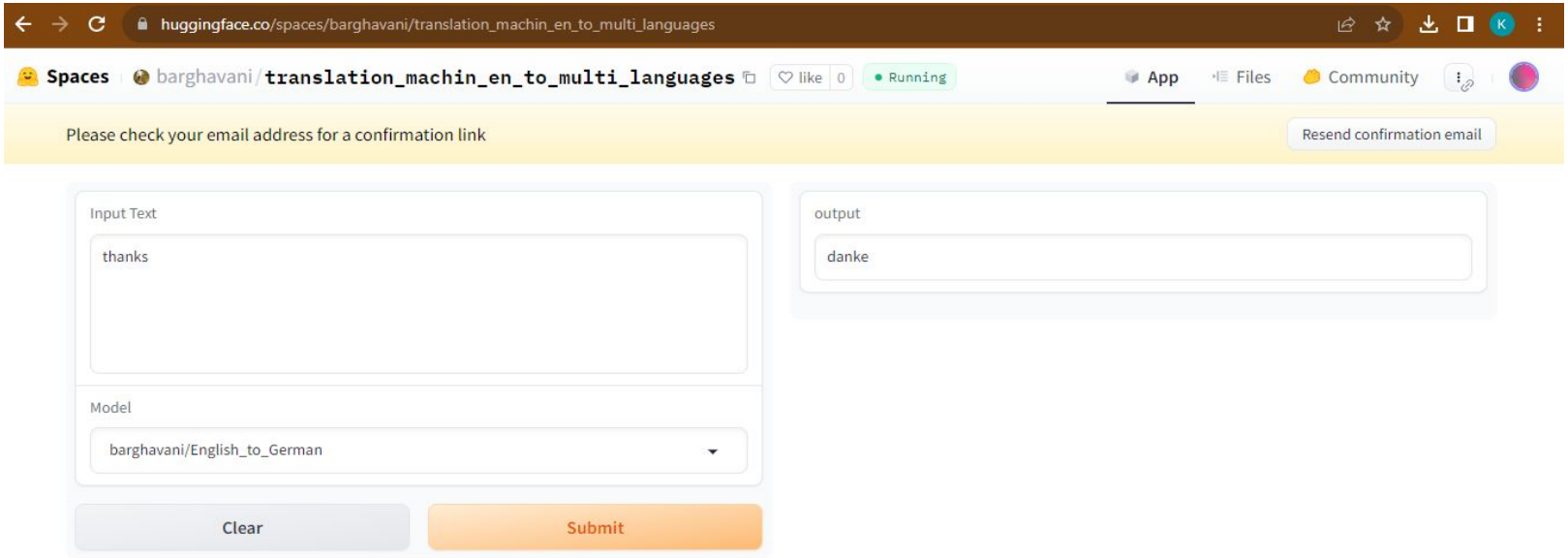
The T5 model has been applied to English, French, Hindi, Romanian, German, and Farsi languages. The model was fine-tuned to improve translation of low-frequency words and create a bilingual lexicon from parallel corpora.

Word Alignments in NMT

Word alignment is crucial for generating bilingual lexicons, understanding the style of webpages, and improving translation of low-frequency words. The attention mechanism was trained as the alignment head, using an unsupervised approach.

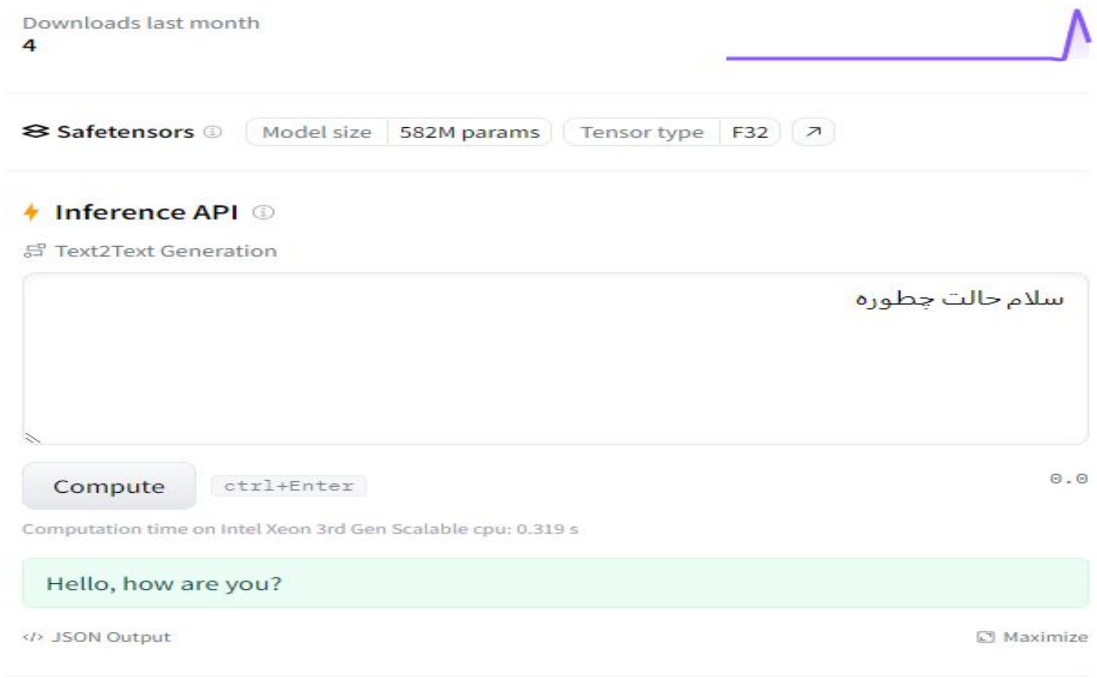
Hugging Face Translations Live Demonstration:

English to German Translation



Hugging Face Translations Live Demonstration:

Farsi to English Translation



Conclusion

Primary results show promising performance for both multi-head self attention with Bi-directional LSTM and T5 models. Future work includes further optimization and application of these models on larger multilingual datasets.

GitHub Repository

https://github.com/Deep-Learning-Project-Fall-23/Deep_learning_project/tree/main

Hugging face

https://huggingface.co/spaces/barghavani/translation_machin_en_to_multi_languages

https://huggingface.co/spaces/barghavani/Translation_Machine_Farsi_to_English\

References

A Convolutional Encoder Model for Neural Machine Translation”https://arxiv.org/abs/1611.02344”

http://tatoeba.org/eng/terms_of_use

http://creativecommons.org/licenses/by/2.0

Attribution: www.manythings.org/anki and tatoeba.org

https://huggingface.co/t5-small

https://huggingface.co/Helsinki-NLP/opus-mt-en-hi

https://huggingface.co/datasets/cfilt/iitb-english-hindi

persiannlp/parsinlu_translation_en_fa

persiannlp/mt5-base-parsinlu-opus-translation_fa_en

https://www.kaggle.com/code/harshjain123/machine-translation-seq2seq-lstms/input