DEEP LEARNING SEMESTER PROJECT

TEAM MEMBERS:
Divya Shivakumar(B21BB011)
Dhyan Yajnik (B21AI054)
Rahul Naik(B21AI009)
Kovidh Pothireddy(B21AI026)

INSTRUCTOR:
Angshuman Paul

PROBLEM STATEMENT:
Deep Learning model for language translation

DATASET
Link: https://huggingface.co/datasets/wmt14/viewer/de-en
The WMT14 dataset is a translation corpus notable for its substantial size, comprising 4.51 million training examples, along with 3,000 instances each for validation and testing.This dataset consists of phrases and their translations, instead of words and their translations, which is important to preserve the context of the sentence, which would even account for the tone of the sentence, any idioms or slangs used. This would help make the translation more accurate and natural for the user to understand.

WHY DEEP LEARNING BASED APPROACH?
Due to two Big Advantages of DL based models over ML based models, which are:

- DL models retain sequential information of text data after conversion to numbers. Example: “This is my house”, “house is This my”.
- DL models automatically generate features. Therefore, even without knowledge of the German language, we can design a model capable of translating any language.

WHAT ARE SOME CURRENT MACHINE TRANSLATION MODELS USED?
DeepL translator, which is a successful startup, proclaiming to have the most accurate and fastest translator in multiple languages, mainly used CNN - Convolutional Neural Network along with some other algorithms to train their data. But they were replaced by more robust models like RNN - Recurrent Neural Networks.

Google translate uses NMT - Neural Machine Translation which is closely related to Seq2Seq for their translation tasks.

There are some more successful models used for machine translation, which include, Word2Vec, LSTM - Long Short Term memory, and more.

WHICH MODEL TO CHOOSE?
CNN is mainly used for image classification, and is easily outperformed by models like RNN.

RNN or recurrent neural networks is a type of neural network constructed and utilized by recursively applying the same set of weights to a structured input. This process generates a structured prediction for a given input structure or produces a scalar prediction for the input. This is achieved by guiding the machine to traverse the provided structure in a methodical topological order. During text translation, the network basically stores important information from words and sentences that were fed to it earlier to understand the new phrases and words.

Advantages of using a RNN based model are:

- The neural network can handle inputs of varying lengths.
- The computation at step "t" theoretically incorporates information from numerous preceding steps.
- The same set of weights is applied at every timestep, creating symmetry in the processing of inputs.

The issues with RNN based models are

- Vanishing Gradients: The gradients diminish to a point where parameter updates become negligible. This challenge complicates the training of lengthy data sequences.
- Exploding Gradients: This issue occurs when substantial error gradients accumulate, resulting in excessively large updates to the model weights of the neural network throughout the training process.
- Recurrent computation tends to be sluggish, and in practical terms, retrieving information from numerous steps in the past poses a challenge.

These are the major issues that we face while using RNN. So the next model that we considered was LSTM to overcome the issue of vanishing and exploding gradients to an extent.

LSTM(long short term memory) is an upgraded version of RNN which overcomes the issues that we faced while working with the RNN model. But it has certain drawbacks as well

- Though it solves the issue of vanishing/exploding gradient to large extent, it doesn’t remove it completely
- We require a lot of data and time to train an LSTM model
- Prone to overfitting

Transformers have revolutionized the field of natural language processing, offering significant improvements over traditional recurrent neural network (RNN) architectures. Unlike RNNs, which process input sequences sequentially, transformers employ a self-attention mechanism to capture dependencies between words in a sentence simultaneously. This parallel processing capability allows transformers to handle inputs of varying lengths more efficiently and reduces the risk of vanishing or exploding gradients.

Advantages of using transformer-based models include:

- Parallel Processing: Transformers can process input sequences in parallel, making them more efficient than RNNs, especially for longer sequences.
- Attention Mechanism: The self-attention mechanism enables transformers to weigh the importance of different words in a sentence, facilitating better context understanding and more accurate predictions.
- Scalability: Transformers exhibit impressive scalability, capable of handling large datasets and complex tasks with ease.
- Reduced Long-Term Dependency Issues: Unlike RNNs, transformers do not suffer from the vanishing gradient problem, allowing them to capture long-range dependencies more effectively.

EVALUATION METRIC

The BLEU (Bilingual Evaluation Understudy) score is a metric used to measure the quality of machine-translated text by comparing it with reference translations. It quantifies the similarity between the machine-generated translation and the reference translations using n-gram precision and brevity penalty. BLEU scores range from 0 to 1, with higher scores indicating better translation quality. Although widely used for its simplicity and computational efficiency.