NLP Lab Report: Automatic English-French Translation

Work prepared by: Mokadem Adel Abderlkader, Benyamina Yacine Lazreg

1 Introduction

Machine translation has been a longstanding challenge in the field of natural language processing (NLP). The ability to accurately translate between languages is crucial for facilitating effective communication across linguistic barriers. Traditional statistical machine translation approaches, while effective to some extent, often struggled with capturing the intricate nuances and complexities of human languages.

In this lab report, we will delve deep into the mechanism of Language Translation using. We will utilize the robust capabilities of Transformers for English to French translation.

2 Dataset

For this English-French language translation task, we utilized the "**Language Translation (English-French)**" dataset obtained from Kaggle. This dataset is a collection of parallel text data, consisting of sentences in English and their corresponding French translations.

The dataset comprises a total of **175621** sentence pairs, providing a substantial amount of training data for the machine translation model. The sentences cover a wide range of topics and domains, including news articles, literary works, and general conversations, ensuring a diverse representation of language usage.

Below are some sample from "Language Translation (English-French)" dataset:

	English words/sentences	French words/sentences
151138	You can't put toothpaste back in the tube.	Tu ne peux pas remettre le dentifrice dans le ...
3088	You're fair.	Tu es juste.
124027	Have you looked at these brochures?	Avez-vous regardé ces brochures ?
121974	The parking lot is free of charge.	Le parking est gratuit.
59458	Could we talk in private?	Pourrait-on parler en privé ?
106545	What're you all dressed up for?	Pourquoi es-tu tout endimanché ?
60798	I can't give up my dream.	Je ne peux abandonner mon rêve.
71670	This book costs 3,000 yen.	Ce livre coûte 3000 yens.
136250	Could we have a table near the window?	Pourrions-nous avoir une table près de la fenê...
131248	Things aren't always as they appear.	Les choses ne sont pas toujours ce qu'elles se...

Figure 1: Some sample from the "Language Translation (English-French)" dataset.

3 Architecture

The implementation of a sequence-to-sequence (Seq2Seq) Transformer model for language translation task involves carefully setting up various components and parameters.

First we define the vocabularies for both the source and target languages, which helps in converting words into numerical representations that the model can process. The embedding size is set to 192, meaning each word is represented by a 192-dimensional vector.

The Transformer architecture uses 6 attention heads, which means it looks at different parts of the input sequence simultaneously to understand various aspects of the context. The feedforward network has a dimension of 192, providing the capacity to learn intricate patterns. Both the encoder and decoder have 3 layers each.

The encoder's job is to take the input sentence in the source language and convert it into a series of vectors that capture the meaning and context of the sentence. It does this through multiple layers of self-attention mechanisms and feedforward neural networks. The self-attention mechanism enables the encoder to weigh the importance of each word in the sentence relative to others, helping it to understand the context. For example, in the sentence "The cat sat on the mat," the encoder learns how each word relates to the others, capturing the overall meaning.

The decoder then takes this context-rich representation from the encoder and, along with the target sequence (shifted to ensure it doesn't "see" future words), generates the translation in the target language. The decoder utilizes both self-attention and cross-attention mechanisms. Self-attention allows the decoder to focus on the words generated so far while understanding their relationships. Cross-attention enables the decoder to incorporate information from the encoder's output, ensuring it considers the context of the entire input sequence during the translation process. This step-by-step process continues until the entire translation is generated.

The model contains around 14.5 million parameters.

4 Training Experimentation

The model was trained for 50 epochs on Kaggle's A100 GPU, with CrossEntropy Loss, Adam optimizer, a learning rate of 0.0001 and a batch size of 128. The total training time took around 2h.

5 Results and Discussion

Following the implementation outlined in the experimental section, the following results were obtained:

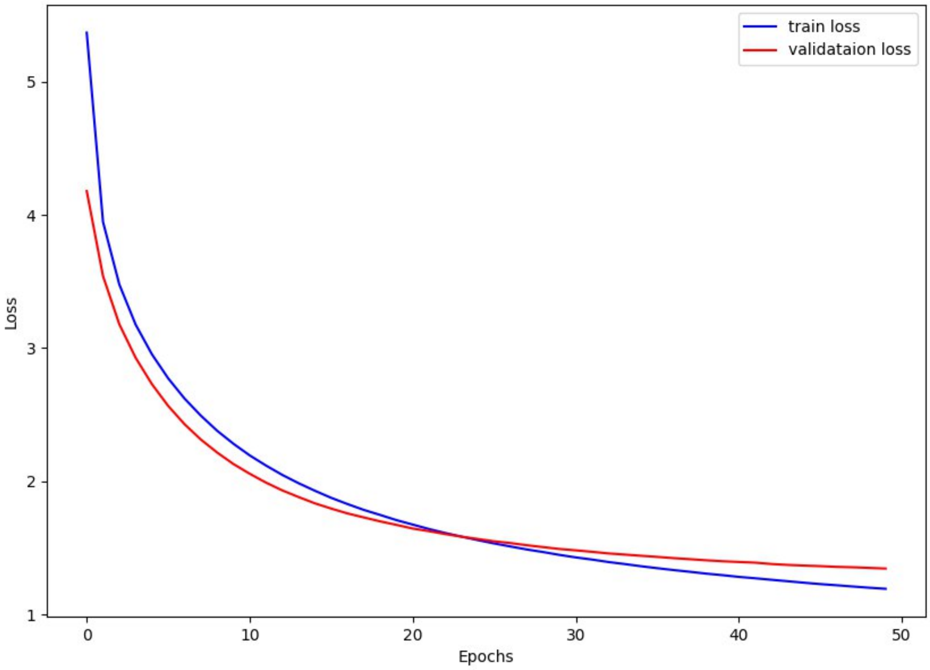

Figure 2: Training and Validation loss graph.

Initially, the training loss was relatively high, reflecting the model's initial state of uncertainty and the complexity of the translation task. However, as the training progressed, the loss steadily decreased, demonstrating the model's ability to capture the intricate patterns and nuances present in the English-French language pair.

To comprehensively assess the performance of the English-French translation model, we employed two widely-used evaluation metrics: BLEU (Bilingual Evaluation Understudy) score and ROUGE (Recall-Oriented Understudy for Gisting Evaluation) score.

5.1 BLEU score (Bilingual Evaluation Understudy)

BLEU (bilingual evaluation understudy) is a metric used to evaluate the generated text of machine translation models or models that are trying to solve similary task ,it was invented in 2001 at IBM,it is known for having a high correlation with human annotations.
The BLUE score compares separate parts of a text with a set of reference texts, and assigns a score to each part then these scores are averaged to give the final score, the higher the score, the better the model.

BLUE-1	BLUE-2	BLUE-3	BLUE-4
0.3257	0.2092	0.1271	0.0736

Table 1: BLEU Score

5.2 ROUGE score (Recall-Oriented Understudy for Gisting Evaluation)

ROUGE, or Recall-Oriented Understudy for Gisting Evaluation, is a metric used to measure the similarity between a machine-generated summary and reference summaries. It does so by comparing overlapping n-grams. ROUGE scores range from 0 to 1, with higher scores indicating a greater similarity between the automatically generated summary and the reference summaries.

	ROUGE-1	ROUGE-2	ROUGE-L
F1	0.2123	0.0590	0.2028
Precision	0.2190	0.0611	0.2091
Recall	0.2147	0.0597	0.2051

Table 2: ROUGE Score

6 Conclusion

In this lab report, we explored the application of the PyTorch Transformer architecture for the task of English-French language translation. While the transformer model demonstrated promising capabilities in capturing long-range dependencies and leveraging self-attention mechanisms, the overall translation performance, as indicated by the evaluation metrics, leaves room for improvement.

The training process was monitored through the loss graph, exhibiting a steady convergence. However, the BLEU and ROUGE scores, which are widely adopted evaluation metrics in machine translation, revealed suboptimal performance in generating high-quality translations that accurately preserve the semantic integrity and essential information of the source sentences.