Project Presentation

CS435/635 Deep Learning

ENGLISH TO HINDI MACHINE TRANSLATION USING TRANSFORMER

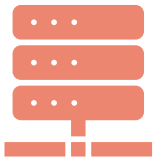
Submitted To-
Dr. Chandresh Kumar Maurya
Course Coordinator

Submitted By-
Kuldeep Pathak
PHD (2301101009)
Amit Singh Yadav
PHD (2301201001)

Table of Contents

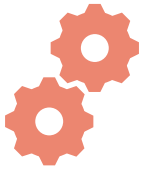
- Introduction
- Motivation
- Objective
- Methodology
- Dataset & Transformer Architecture
- Result
- Conclusion
- References

Introduction



Machine Translation (MT) is the automated process of translating text.

English and Hindi are widely spoken languages with distinct linguistic characteristics.



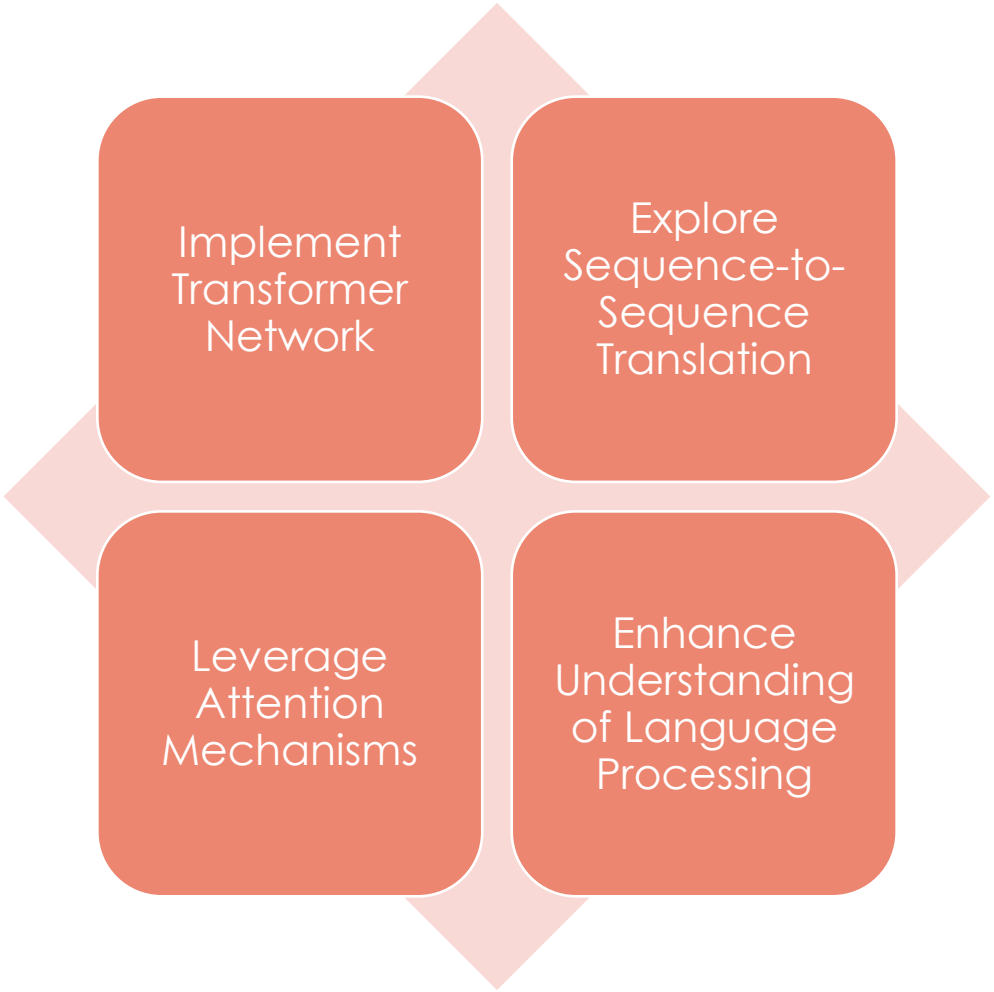
It provides us the opportunity to explore and implement the Transformer architecture.

Transformer model shows promising results compared to traditional statistical machine translation and other neural machine translation models. [3]

Motivation

- Hands-On Learning Opportunity.
- Transformer's Adaptability to Sequences.
- Understanding Attention Mechanisms.
- Challenges of Cross-Linguistic Translation.

Objective


Implement Transformer Network
Explore Sequence-to-Sequence Translation
Leverage Attention Mechanisms
Enhance Understanding of Language Processing

Methodology

Data Preparation

- Used Positional Encoding
- Tokenizer

Model Architecture

- Encoder
- Decoder

Methodology

Training

- Split dataset into 70% Training and 30% for testing

Evaluation

- Observe Training and Testing loss

Optimization and Fine tuning

- Run multiple times by changing the hyperparameter

Dataset

Used a dataset from Kaggle[2].

- English to Hindi Text

The dataset contains more than 130000 records.

- Used only 100000 records.

The dataset is prepared using the corpus and scrapping the web.

The shape of dataset is:

- 130476 Row
- 2 Column

Transformer Components

Embedding Layer

Encoder and Decoder Stacks

Attention Mechanism

Positional Encoding

Feed-Forward Neural Networks

Transformer Architecture

Main Component

- Self-Attention Mechanism
- Multi-Head Attention
- Encoder and Decoder Stacks

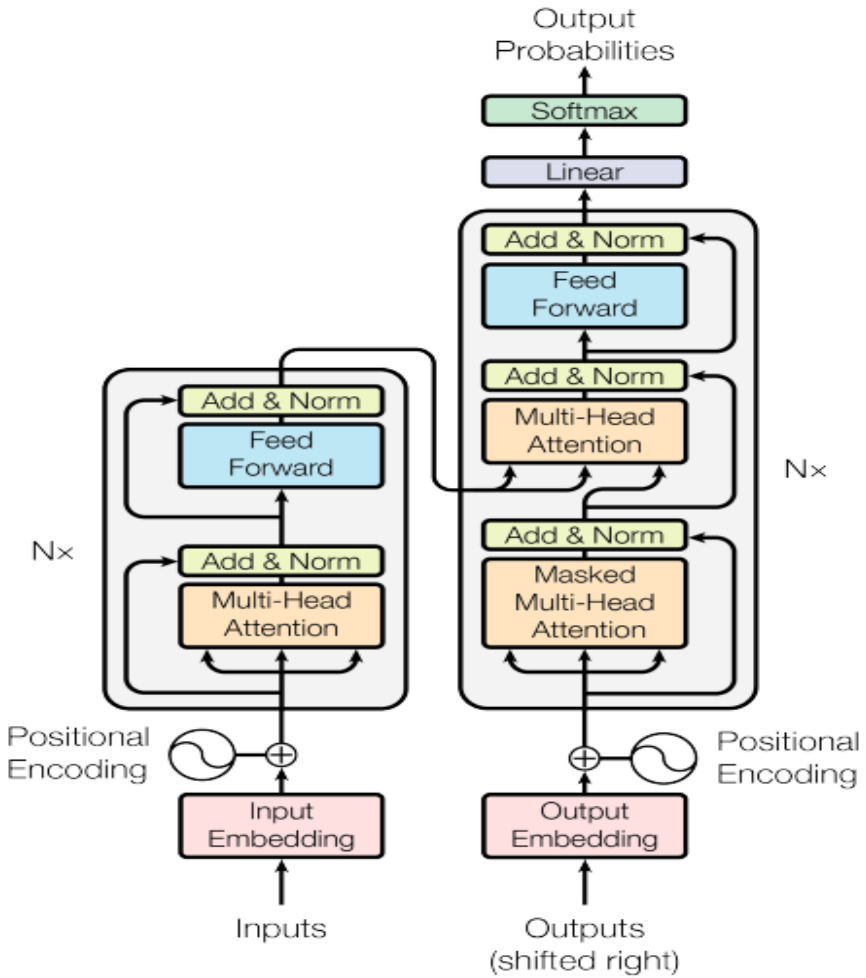


Figure 1. Attention Architecture[1]

Transformer Architecture (contd.)

Main Component

- Positional Encoding
- Feed-Forward Neural Networks

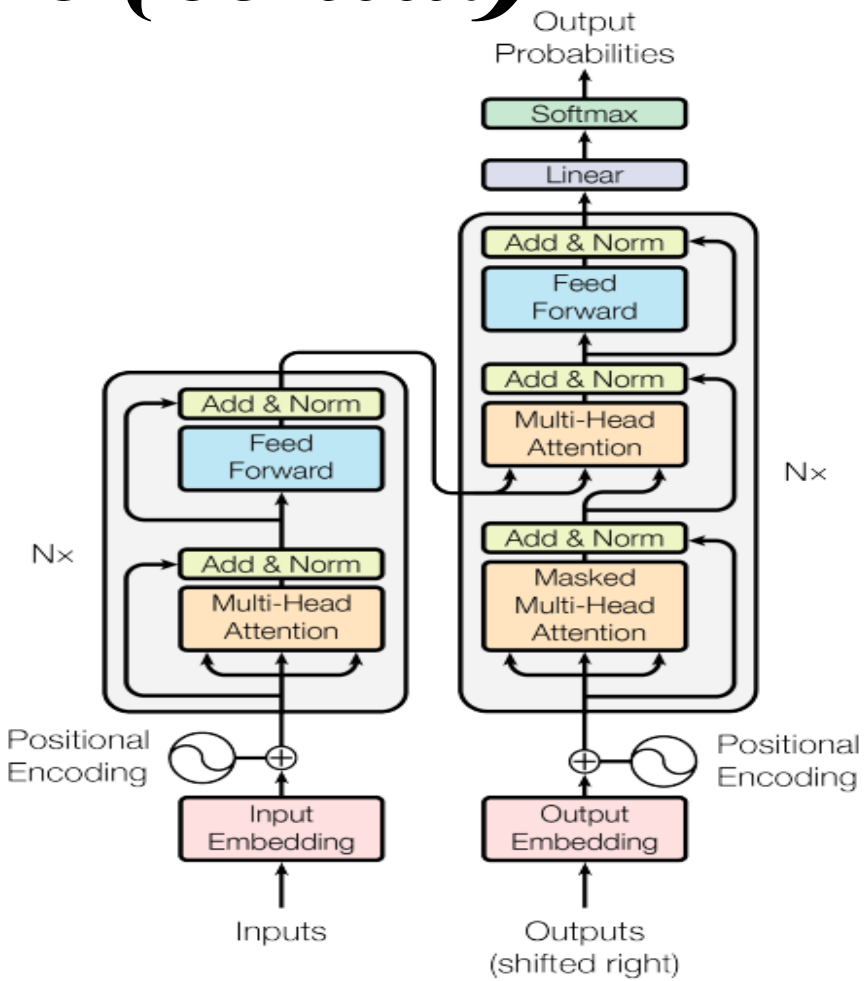


Figure 1. Attention Architecture[1]

Result

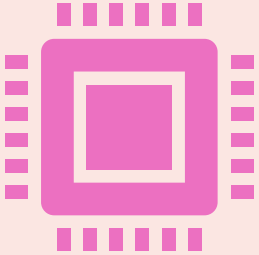
Experimental Setup:

Cross Entropy Loss

Adam optimizer

AutoTokenizer from the pretrained "Helsinki-NLP/opus-mt-en-hi"

P100 GPU



Hyperparameters:

Batch Size: 16

Learning Rate: 0.01

Max Input Length: 512

Max Output Length: 512

Training Epochs: 25

Result

Train Loss: 1.8217

Test Loss: 1.6819

The model demonstrates a decreasing trend in both train and test loss.

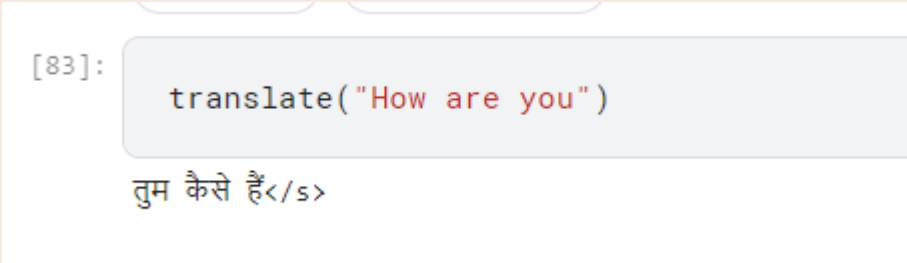
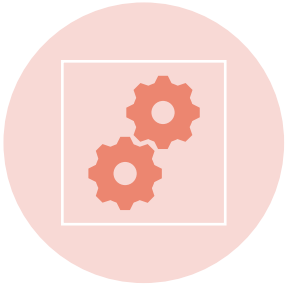


Figure 2. Final Output

Conclusion

Comprehensive Understanding

Effective Utilization



Optimization and Fine-Tuning

Practical Application

References

1. Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).
2. https://www.kaggle.com/datasets/preetviradiya/english-hindi-dataset
3. Watve, Abhinav Y., and Madhuri A. Bhalekar. "English to hindi translation using transformer." 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA). IEEE, 2023.

- Thank You