CSE 676 (Summer 2023): Deep Learning Final Project

Machine Language Translation with DL

Team 13

Sailesh Reddy Sirigireddy, saileshr

Shri Harsha Adapala Thirumala, sadapala

Chandini Kondinolu, ckondino

University at Buffalo The State University of New York

Project Description

The major objective of this project is create a machine translator that takes prompt from language A and translates into language B where A and B are languages we train our model on. Our goals is to have fluency and naturality of the translated texts and the output must be as a sentence generated by a human who is fluent in the language.

We will then compare the accuracies of various architectures in solving the above task of translating from English to French.

Background

- *Sequence to Sequence Learning with Neural Networks by Sutskever et al. (2014)* – Basic Architecture for Seq2Seq Learning with LSTM
- *Neural Machine Translation by Jointly learning to Align and Translate by Bahdanau et al. (2015)* – Introduces Attention Mechanism so as to effectively translate longer sentences
- Effective Approaches to Attention-based Neural Machine Translation by Luong et al. *(2015)* – Different effective ways to introduce attention to the model architecture

Dataset

The data is taken from the parallel language corpus available in Kaggle. DGT-TM is a translation memory (sentences and their manually produced translations) in 24 languages.

We further selected a dataset for English to French translations.

In total we are using around 135,000 sentences of upto 25 word length for training and test.

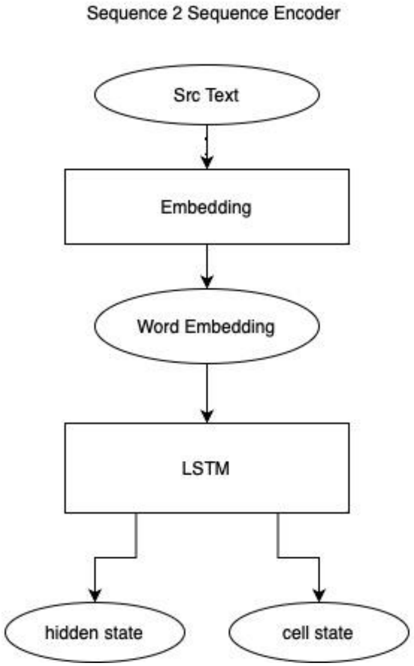
Dataset Preprocessing

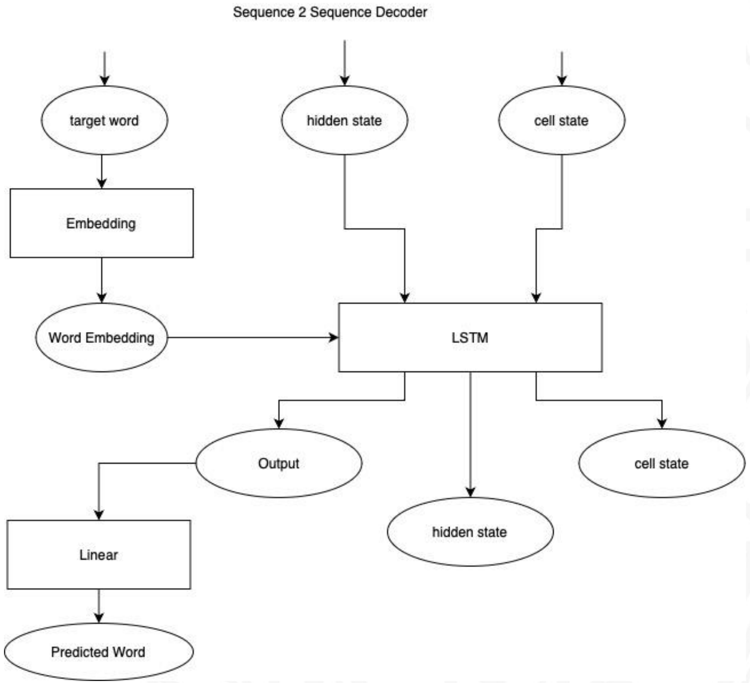
- We used spacy, torchtext and NLTK libraries for preprocessing
- Before tokenization we converted sentences to lower case and then removed special characters and digits
- We then tokezined and added <sos> and <eos> tags at the start and end.
- We then built vocabularies for English and French feeding the training data to Spacy models.

Dataset Preprocessing - Challenges

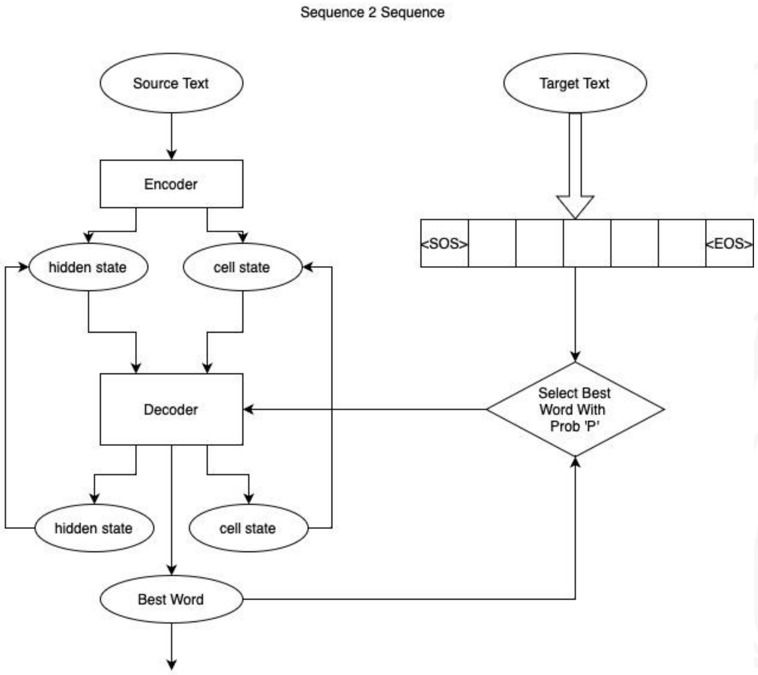
- Excessive Padding:
  - If we padding the entire corpus with maximum sentence length, training might not happen.
  - To overcome this we are using sorting and batch padding techniques.
  - It also improves training time and decreasing the computational expenses.
- Word Embedding:
  - We are using a custom Embedding layer to do this.

Vanilla Seq2Seq DataFlow

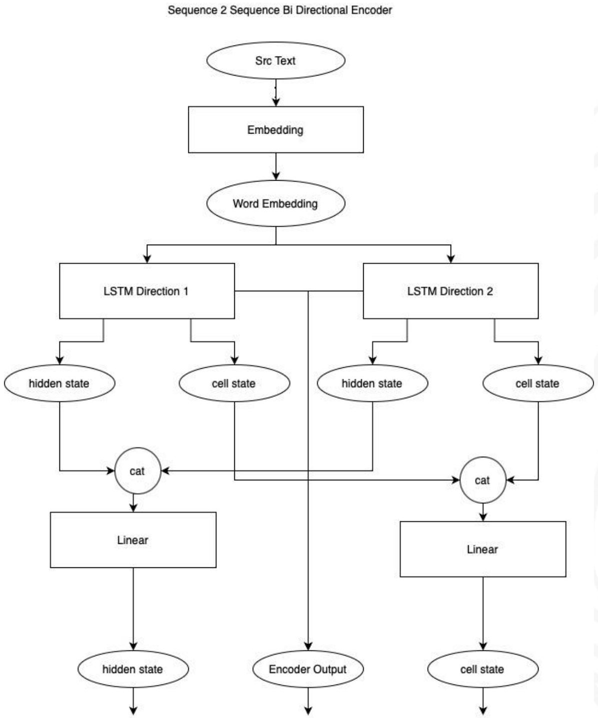




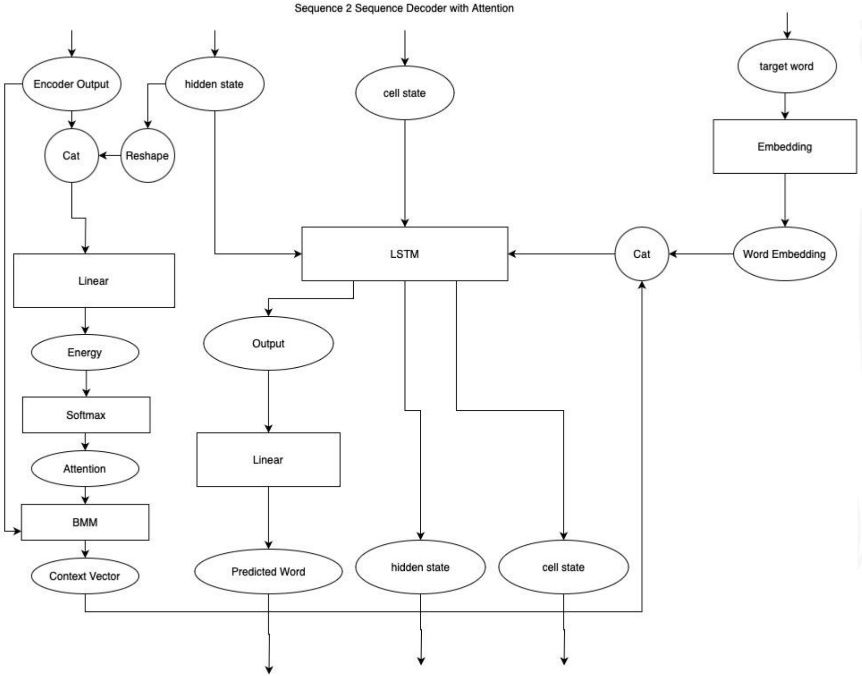
Vanilla Seq2Seq DataFlow



Bidirectional Encoding DataFlow

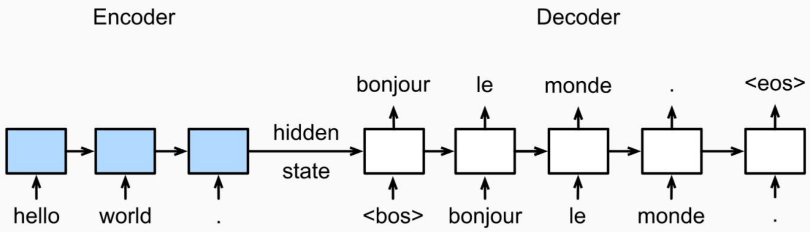


Seq2Seq with Attention DataFlow



Implementation - Vanilla Seq2Seq model

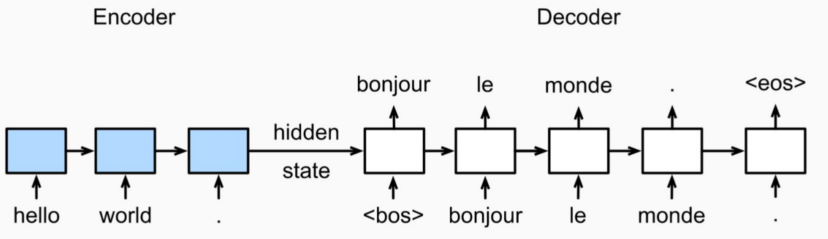
- Hyperparameters
  - Learning Rate = 3e-4
  - Number of epochs = 50
  - Batch Size = 64
  - Encoder and Decoder Dropout = 0.5
  - Encoder and Decoder Embedding Size = 300
  - Teacher Force Ratio = 0.6
  - LSTM Number of layers = 2
  - Optimizer = Adam



The sequence to sequence model architecture.

Implementation - Seq2Seq with Bidirectional LSTM

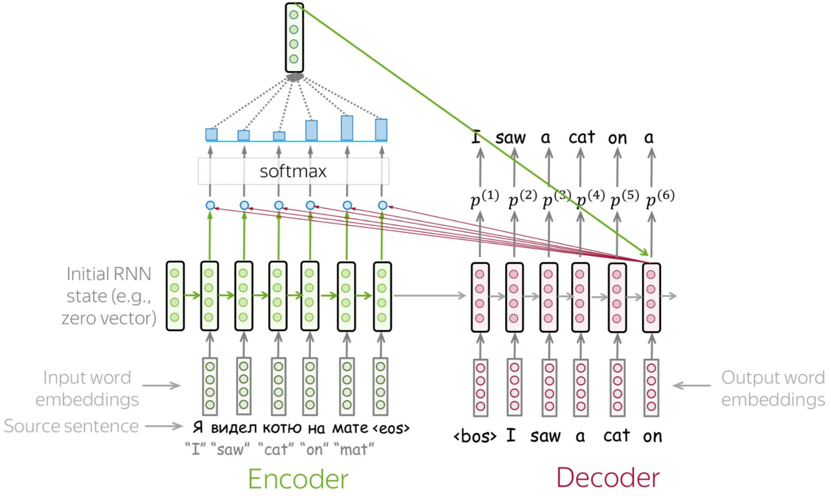
- Hyperparameters
  - Learning Rate = 3e-4
  - Number of epochs = 50
  - Batch Size = 64
  - Encoder and Decoder Dropout = 0.5
  - Encoder and Decoder Embedding Size = 300
  - Teacher Force Ratio = 0.6
  - LSTM Number of layers = 2
  - Optimizer = Adam



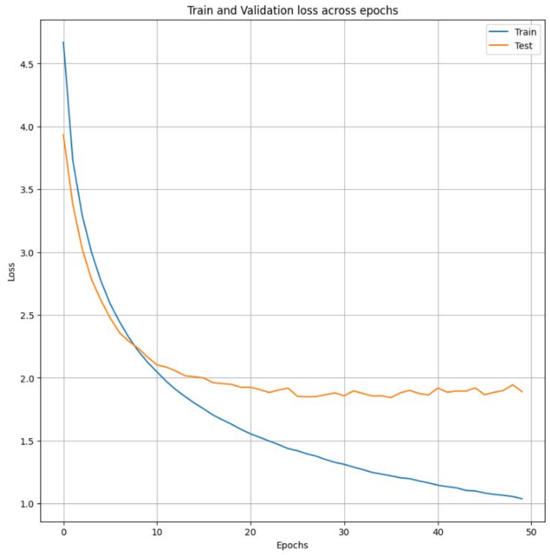
The sequence to sequence model architecture.

Implementation - Seq2Seq Model with Attention

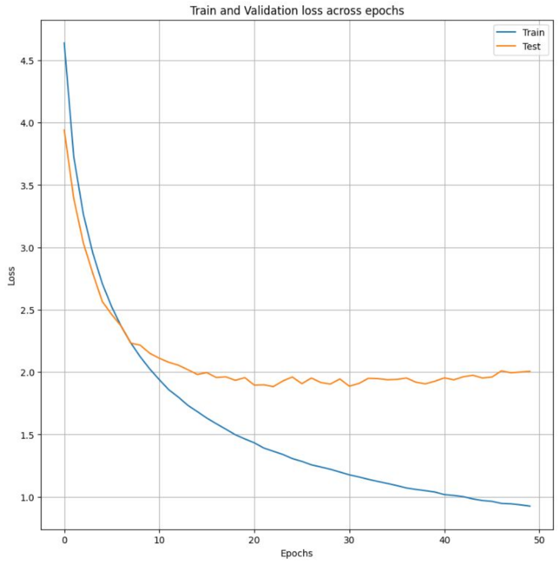
- Hyperparameters
  - Learning Rate = 3e-4
  - Number of epochs = 50
  - Batch Size = 64
  - Encoder and Decoder Dropout = 0.6
  - Encoder and Decoder Embedding Size = 300
  - Teacher Force Ratio = 0.6
  - LSTM Number of layers = 1
  - Optimizer = Adam



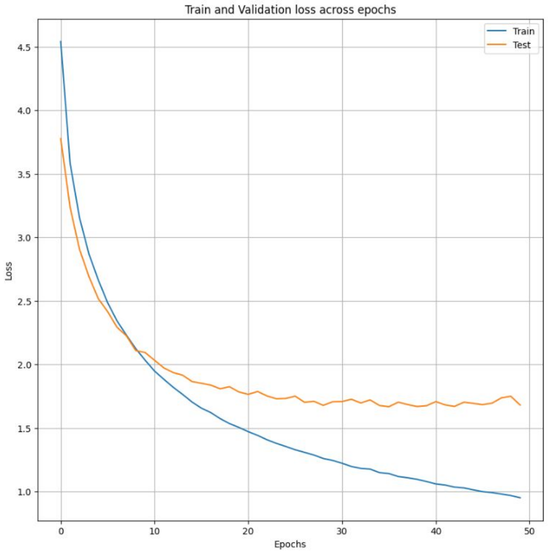
Results - Loss Plots



Seq2Seq

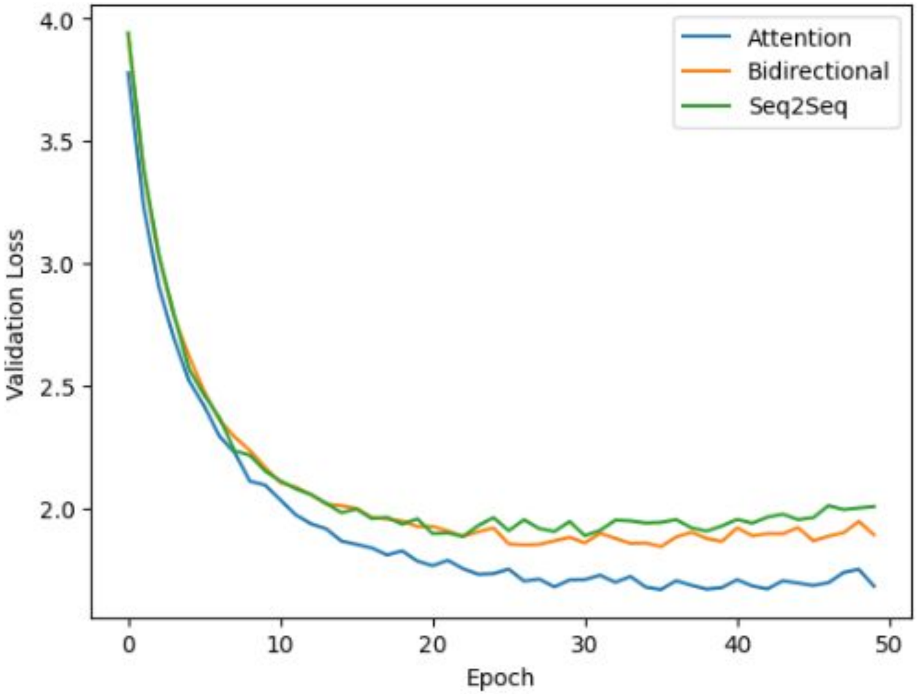


Bidirectional



Seq2Seq with Attention

Results - Validation Loss Comparison



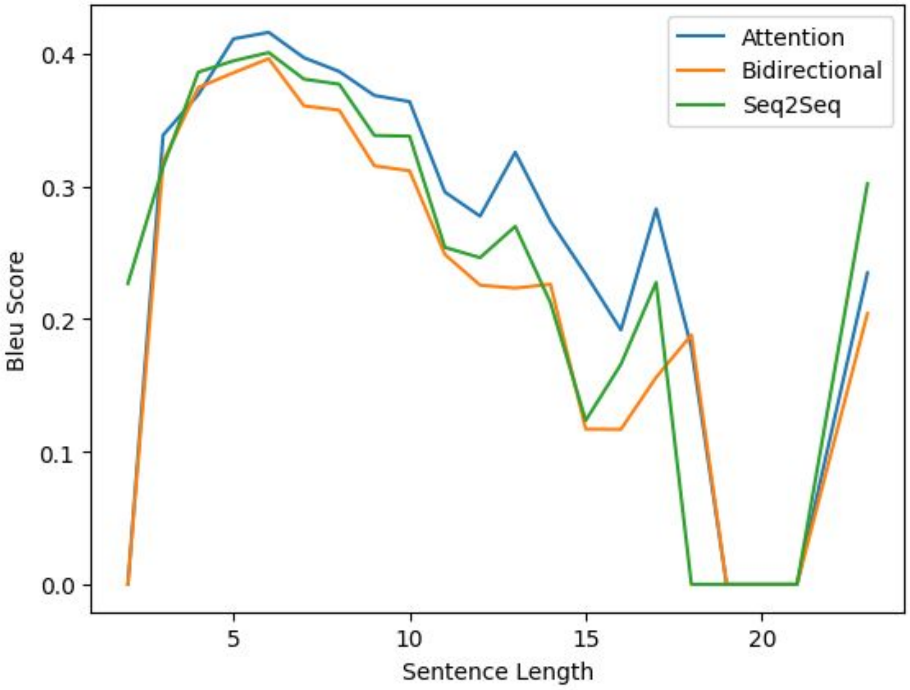
BLeU(BiLingual Evaluation Understudy) Score

- BLeU Score is a metric for automatically evaluating machine-translated text. It is a number between zero and one that measures the similarity of the machine-translated text to a set of high quality reference translations.
- A value of 0 means that the machine-translated output has no overlap with the reference translation (low quality) while a value of 1 means there is perfect overlap with the reference translations (high quality).

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 40	Understandable to good translations
40 - 50	High quality translations
50 - 60	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

Evaluation - BLeU Score Comparison

- Seq2Seq - 0.365
- Bidirectional - 0.369
- Seq2Seq with Attention - 0.406



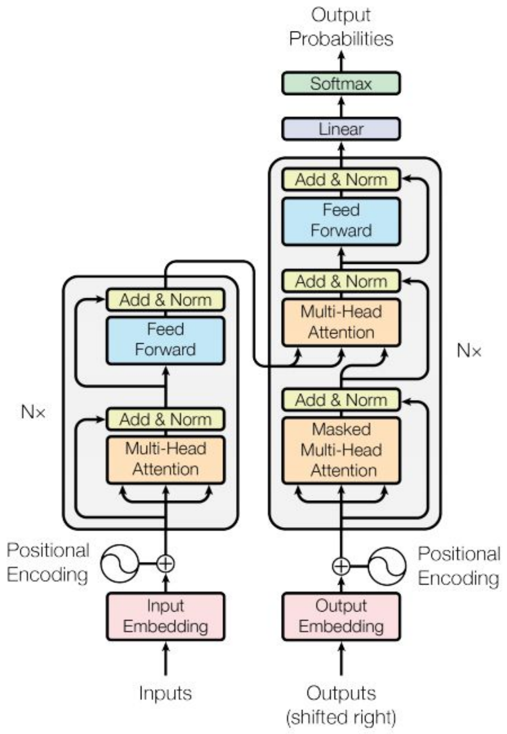
Demo

Key Takeaways

- We observed we had slightly better loss values for Seq2Seq with Bidirectional Encoding over Vanilla Seq2Seq. And Seq2Seq with Attention had significantly better loss value than other two models.
- We also observed that Seq2Seq with Attention had an better overall BLeU Score as well as slightly better BleU score over longer sentences.
- Hence, Attention is all you need.

Further Improvements

- Transformer based Architecture: Uses self attention mechanism to capture dependencies between words in a sentence effectively
- Larger dataset with longer sentences
- Data Augmentation: Use techniques such as word swapping, sentence shuffling etc.



Contribution

Name	Part	Contribution
Sailesh Reddy Sirigireddy	Vanilla Seq2Seq	35%
Chandini Kondinolu	Bidirectional Encoding	30%
Shri Harsha Adapala Thirumala	Seq2Seq with Attention	35%

Thank You