
SEQ2SEQ MODEL FOR MACHINE TRANSLATION

AI6127: Deep Learning for Natural Language Processing

Riemer van der Vliet
G2304212K

Submission Date: 23 April 2024

Abstract

This report investigates Seq2Seq models for machine translation, analyzing their performance across various architectures. Beginning with a baseline GRU model, we explore LSTM, Bi-LSTM, attention mechanisms, and transformer encoders. Using Rouge scores, we assess translation quality. Results reveal attention-based models' superiority, particularly in bi-gram matching, showcasing their adeptness at capturing contextual nuances. Notably, Transformer models show promise but lag in convergence after two epochs, indicating potential for further improvement through extended training and hyperparameter optimization. This underscores attention mechanisms' pivotal role in advancing Seq2Seq models for translation tasks, highlighting refinement to achieve optimal performance.

1 Introduction

This report focuses on the implementation and comparative analysis of various Seq2Seq models for the task of machine translation. We begin with a baseline model utilizing GRU for both the encoder and decoder components, renowned for their efficiency and capacity to mitigate vanishing gradient problems common in traditional RNNs. Subsequently, we explore the substitution of GRU with LSTM in the encoder and decoder, aiming to leverage LSTM's more complex gating mechanisms for better information flow. Additionally, the implementation includes a variant with a bidirectional LSTM (Bi-LSTM) in the encoder, designed to enhance the model's context understanding by processing the input sequence in both forward and reverse directions.

2 Methodology

In this report the following Seq2Seq models are implemented and compared for machine translation:

- Original GRU-based model.
- LSTM replacement for GRU.
- Bi-LSTM in the Encoder.
- Addition of the attention mechanism.
- Transformer Encoder integration.

3 Experiments

3.1 Experimental Setup

3.1.1 Dataset

This dataset serves as a practical foundation for implementing and testing seq2seq models, specifically tailored towards converting sentences from English to French. The primary

source of this data is Tatoeba.org, a community-driven project aimed at providing accessible translations for various sentences across numerous languages. The Tatoeba project’s extensive database allows for a broad range of applications, from language learning tools to advanced research in machine translation.

For this project, the English-French translation pairs are available through an ancillary resource, <http://www.manythings.org/anki/>. This facilitates the direct application of these pairs in training seq2seq models by providing a simple, structured format. Each line in the dataset file contains an English sentence followed by its French translation, separated by a tab character, making it straightforward to parse and use in training algorithms.

“I am cold. J’ai froid.”

The datasets are preprocessed to remove special characters, convert text to lowercase, and tokenize the sentences. The tokenization process involves splitting the sentences into individual words, which are then used to create a vocabulary for the model. The vocabulary is a set of unique words present in the dataset, which serves as the basis for encoding and decoding text sequences during training and inference.

3.1.2 Rouge Scores

ROUGE scores, an acronym for Recall-Oriented Understudy for Gisting Evaluation, are a set of metrics used to evaluate the quality of text generated by a model. The ROUGE scores are based on the comparison of n-grams between the generated text and the reference text. A higher ROUGE score indicates a better match between the generated text and the reference text.

The difference between rouge1 and rouge2 is the number of words in the n-gram. Rouge1 compares unigrams, while Rouge2 compares bigrams. The fmeasure, precision, and recall are the three metrics used to evaluate the quality of the generated text.

3.2 Results

Here the results are discussed for the methods mentioned in Methodology. **Baseline GRU Model** This Gated Recurrent Unit (GRU) model serves as the baseline for comparison with other models. GRUs are known for their efficiency and ability to mitigate vanishing gradient problems common in traditional RNNs.

LSTM Model This is particularly useful in tasks where the past context is important for predicting the future, such as time series forecasting or unidirectional language translation

Bi-LSTM Model This approach allows the network to have both forward-looking (future) and backward-looking (past) context at every point in the sequence

Model with Attention Attention allows the model to focus on different parts of the input sequence for each step of the output sequence, providing a way to dynamically weigh the input’s relevance. This is especially useful in translation where the importance of input words can vary depending on the context of the output being generated.

Transformer Encoder Model

GRUs process data sequentially, which inherently limits the speed of computation, especially for long sequences. The Transformer model, addresses sequential processing limitations by employing self-attention mechanisms, enabling parallel computation and facilitating more effective learning across long sequences.

Model	Rouge1			Rouge2		
	Fmeasure	Precision	Recall	Fmeasure	Precision	Recall
Baseline GRU	0.6239	0.5904	0.6722	0.4336	0.4035	0.4785
LSTM	0.6016	0.5764	0.6378	0.4040	0.3818	0.4378
Bi-LSTM	0.6204	0.5912	0.6619	0.4232	0.3964	0.4631
Attention model	0.6881	0.6508	0.7393	0.5141	0.4772	0.5674
Transformer model	0.2069	0.2604	0.1793	0.1209	0.1615	0.1029

Table 1: Evaluation scores - Rouge metrics for the different models

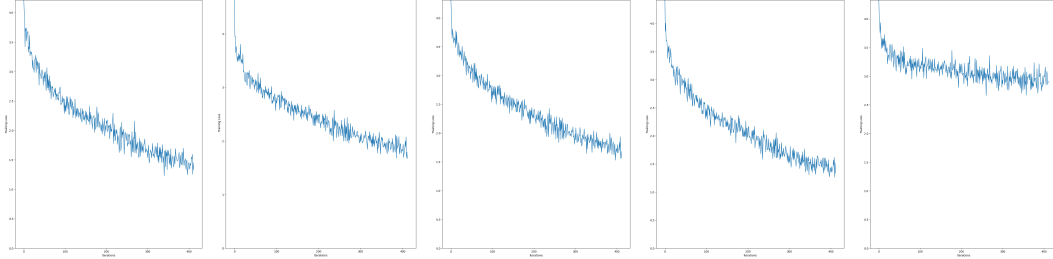


Figure 1: GRU Figure 2: LSTM Figure 3: Bi-LSTM Figure 4: Attention Figure 5: Transformer

Figure 6: Comparison of Seq2Seq Models training over 2 epochs.

4 Discussion

As you can see in table 1, the model with attention outperforms the other models in terms of Rouge scores. This is expected as the attention mechanism allows the model to focus on different parts of the input sequence for each step of the output sequence, providing a way to dynamically weigh the input’s relevance. This is especially useful in translation where the importance of input words can vary depending on the context of the output being generated.

It’s worth noting that the attention model scores particularly high on the Rouge2 set, which implies that attention has some additional benefits when it comes to bi gram matching. That none of the other models possess.

When it comes to learning there should be considered hyperparameter optimization, but as can be seen in 1, 2, 3, 4, and 5 the models are learning at a similar pace and all converging. However, the transformer model, after 2 epochs, seems to be much further removed from converging. Which implies more epochs would benefit this model in particular.

5 Conclusion

In conclusion, the comparative analysis of Seq2Seq models for machine translation reveals that the model incorporating attention mechanism outperforms others in terms of Rouge scores. This superiority is attributed to the attention mechanism’s capability to dynamically weigh the relevance of input words based on the context of the output being generated. Particularly, the attention model excels in bi-gram matching, which is evident from its high Rouge2 score compared to other models.

Moreover, while all models exhibit similar learning trends and convergence patterns, the transformer model appears to lag in convergence after 2 epochs. This suggests that further

training epochs would likely benefit the transformer model, emphasizing the importance of hyperparameter optimization for optimal performance across all models.

In essence, the study underscores the significance of attention mechanism in enhancing the performance of Seq2Seq models for machine translation tasks, and it highlights the potential for further improvements through continued training and fine-tuning of hyperparameters.

6 References

7 Appendix

Codebase can be found under ‘./src’ in the git repository <https://github.com/Riemer1818/AI6127-dNLP-Ass2>.