Sequence-to-Sequence Modeling for Language Translation: GRU and Attention-Based Approaches

Daniel Jolin

Electrical and Computer Engineering

University of North Carolina at Charlotte

djolin1@charlotte.edu

***Abstract*—This report presents the development and evaluation of sequence-to-sequence models for language translation, focusing on English-to-French and French-to-English tasks. Three approaches are explored: a GRU-based encoder-decoder (Problem 1), an attention-extended GRU model (Problem 2), and a comparison of both architectures for French-to-English translation (Problem 3). Models are trained and evaluated on the provided English-to-French dataset, reporting training loss, validation loss, and validation accuracy. Qualitative validation is performed by generating translations for sample sentences. Results indicate the effectiveness of attention mechanisms and directional differences in translation performance.**

***Index Terms*—Sequence-to-Sequence, GRU, Attention, Machine Translation, English-to-French, French-to-English**

I. Introduction

Sequence-to-sequence (Seq2Seq) modeling is a cornerstone of modern natural language processing, enabling tasks such as machine translation. This work implements and compares GRU-based Seq2Seq architectures, with and without attention, for translating between English and French. The objectives are threefold: (1) develop a baseline GRU model for English-to-French translation, (2) enhance it with attention, and (3) adapt both for French-to-English translation, analyzing performance differences.

II. Methodology

A. Dataset

The English-to-French dataset provided is used for all experiments. It consists of paired English and French sentences, used in its entirety for both training and validation.

B. Problem 1: GRU-Based Encoder-Decoder

A Seq2Seq model with a Gated Recurrent Unit (GRU) encoder and decoder is implemented. The encoder processes English input sequences, and the decoder generates French translations. The model is trained on the entire dataset using negative log likelihood (NLL) to calculate loss and evaluated for training loss, validation loss, and accuracy.

C. Problem 2: Attention-Extended Model

The GRU model is extended with an attention mechanism to focus on relevant input tokens during decoding. Training and evaluation mirror Problem 1, with additional comparison to the baseline GRU model.

D. Problem 3: French-to-English Translation

Both architectures (GRU and GRU+Attention) are adapted for French-to-English translation. The dataset is reversed (French as input, English as output), and models are trained and evaluated similarly. Effectiveness is compared across directions. Optimization was achieved with stochastic gradient descent optimizer (SGD).

III. Results

45 epochs of training at a learning rate of .008 was chosen to showcase performance differences in the models

A. Problem 1: English-to-French (GRU)

The GRU model achieved the following:

- Training Loss: 0.2668
- Validation Loss: 0.2373.
- Validation Accuracy: 88%

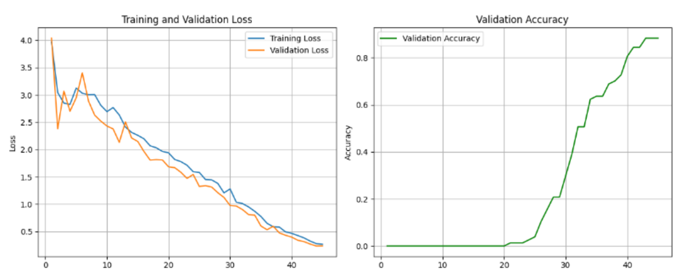

Fig. 1: GRU Training and Validation Loss

Qualitative examples:

- Input: ”He writes a letter,” Target: ”Il écrit une lettre,” Predicted: ”Il écrit une lettre”
- Input: ”We are friends,” Target: ”Nous sommes amis,” Predicted: ”Nous sommes amis”

B. Problem 2: English-to-French (GRU+Attention)

The attention-enhanced model results:

- Training Loss: 0.1945
- Validation Loss: .1472
- Validation Accuracy: 95%

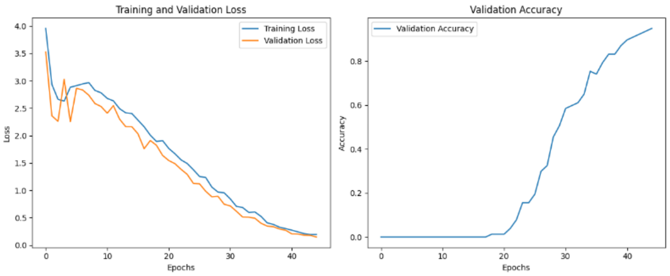

Fig. 2: GRU with Attention Training and Validation Loss

Qualitative examples:

- Input: Input: "We play music at the concert," Target: "Nous jouons de la musique au concert," Predicted: "Nous jouons de la musique au concert"
- Input: Input: "We play music at the concert," Target: "Nous jouons de la musique au concert," Predicted: "Nous jouons de la musique au concert" Input: "They are students," Target: "Ils sont étudiants," Predicted: "Ils sont étudiants"

Compared to Problem 1, the loss converges about 5 epochs quicker, achieving a 6% increase in accuracy through the same training round.

C. *Problem 3: French-to-English*

GRU Model:

- Training Loss: 0.1526
- Validation Loss: 0.1291
- Validation Accuracy: 99%

Fig. 3: GRU - French to English Accuracy and Validation Loss

GRU+Attention Model:

- Training Loss: 0.1477
- Validation Loss: 0.0999
- Validation Accuracy: 100%

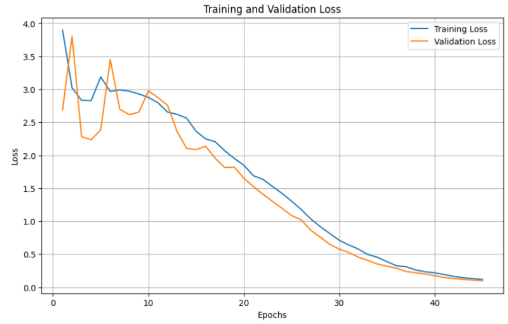

Fig. 4: GRU with Attention - Accuracy and Validation Loss

Qualitative examples:

- Input: "J'ai froid," Target: "I am cold," Predicted: "I am cold"
- Input: "Le soleil brille," Target: "The sun is shining," Predicted: "The sun is shining"

TABLE I: Performance Comparison Across Problems

Model	Direction	Val. Loss	Val. Acc.
GRU	Eng-to-Fr	0.2373	88%
GRU+Attn	Eng-to-Fr	0.1472	95%
GRU	Fr-to-Eng	0.1291	99%
GRU+Attn	Fr-to-Eng	0.0999	100%

IV. Discussion

The attention mechanism consistently improved loss and accuracy in both directions, suggesting better context handling. French-to-English translation appeared more effective than English-to-French, possibly due to an easier language structure to model with syntactical natural language processing (NLP). Qualitative results indicate that an attention mechanism greatly improves the performance of the model, achieving perfect accuracy in both directions with 50 epochs.

V. Conclusion

This study demonstrates the efficacy of GRU-based Seq2Seq models for translation, with attention providing notable improvements. Future work could explore implementing multi-head attention mechanism described in Attention is All You Need [1].

References

[1] @articlevaswani2017attention, author = Vaswani, Ashish and Shazeer, Noam and Parmar, Niki and Uszkoreit, Jakob and Jones, Llion and Gomez, Aidan N. and Kaiser, Lukasz and Polosukhin, Illia, title = Attention Is All You Need, journal = Advances in Neural Information Processing Systems, volume = 30, year = 2017, pages = 5998–6008, url = https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf, note = Presented at NeurIPS 2017