

CS572 Assignment 3 Report

Austin Brown

March 30, 2024

1 Introduction

In this assignment, we are looking to improve a Seq2Seq model that translates German to English.

2 Changes

We made the following changes to try to improve our model: using GloVe embeddings, creating a larger training set using back translation, as well as using regularization in the form of weight decay.

3 GloVe Embeddings

By leveraging GloVe embeddings, we aimed to enhance the representation of words in our Seq2Seq model. These embeddings provide a rich semantic context that helps the model better understand the relationships between words in both source (German) and target (English) languages. This can lead to more accurate translations and improved overall performance.

The specific GloVe model used in our project was GloVe 840B 300, which indicates it was trained on a corpus with 840 billion tokens and has an embedding dimensionality of 300.

We did not use GloVe embeddings in our model for back-translation as it increased the time for the model to run while not increasing performance.

4 Back Translation

Because our initial German to English dataset was not very large, we set out to use back translation to generate new data for us to use. We took an English sentence corpus from the WMT 2014 English-German dataset. We did not use the German dataset at all in our training. Because this dataset was so much larger than our original one, we only used 30,000 sentences from it, so that our final dataset would be half our original sentences and half our back translated sentences. The model that we used for back translation was the base Seq2Seq attention model. We got poor results for Back Translation. My current theory is that the sentences we used in Back Translation were much more long and complex than the sentences in the original training set so the model struggled to handle the longer sentences. In the future I would try to cut down the sentences or find a better dataset.

5 Regularization

We also introduced weight decay with a value of $1e-3$ to try and prevent overfitting and make the model generalize better to the hidden test set on Gradescope. This led to us finding our best model.

6 Results

Table 1: Summary of Experimental Results (Attention used on all

Model	Validation
Beam	38.271434226348816
Beam, Regularization	39.43393944817083
Glove, Beam	37.19665462860924
Glove, Regularization, Back-translation	5.100563211857484