DL Assignment 1 Part2

Utkarsh Sharma

April 2024

https://drive.google.com/drive/folders/1HJG2C-feKhI8Q_qFbaC_DQw09aq1TQXm?usp=drive_link

1 Introduction

The aim of this assignment is to develop a model to automatically convert a piece of text that constitutes a mathematical "word problem" into an executable program. To see an example, refer to figure 1.
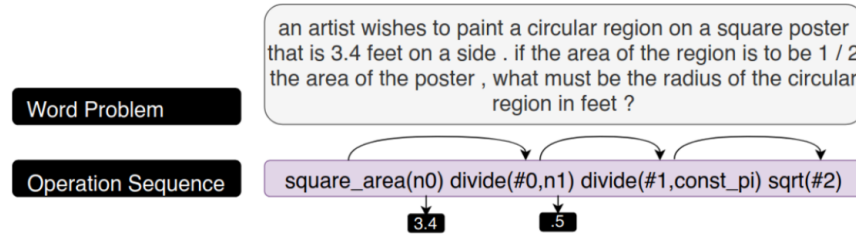

Figure 1: An example word problem converted to an executable program

2 Text to Math program

2.1 LSTM LSTM

Results

The results mentioned below are obtained using a beam size of 10.

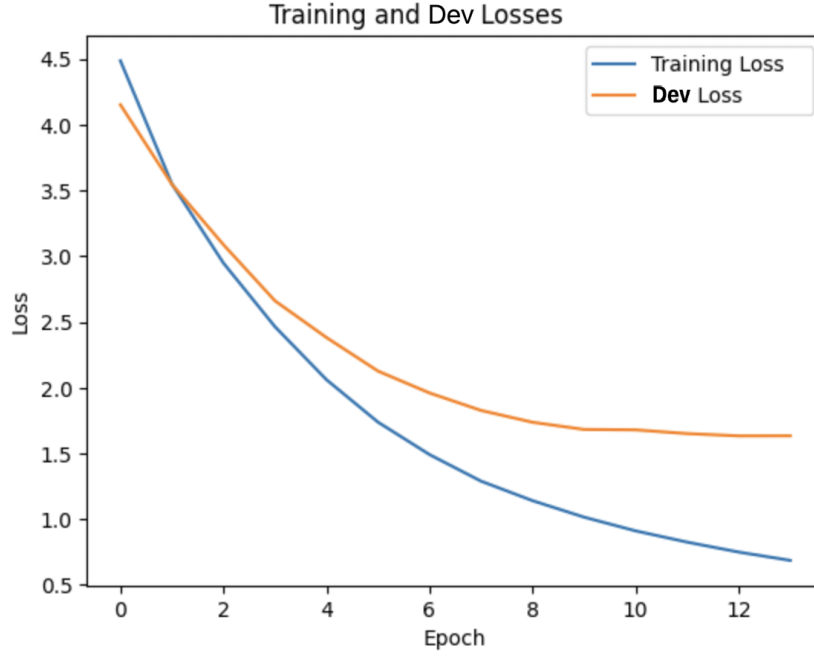| Set	Exact Match	Execution
dev	0.718	0.753
test	0.729	0.771

Figure 2: Losses for LSTM LSTM model v.s. Epochs

2.2   LSTM LSTM + Att

Results

The results mentioned below are obtained using a beam size of 10.

Set	Exact Match	Execution
dev	0.638	0.673
test	0.649	0.681

2.3   Effect of Teacher Forcing Probability

Results

Below, we present a complete picture of the results obtained during the training.

A few observations

When the teacher forcing ratio is increased, we are better fitting to the training data, hence the training loss should decrease, which is what is supported by our data.

This in turn affects the performance on the test data set, as the model probably overfits to the training data on increasing the ratio too much.

Teacher Forcing Ratio	Set	Exact Match	Execution
0.3	dev	0.657	0.687
	test	0.655	0.682
0.6	dev	0.635	0.675
	test	0.645	0.687
0.9	dev	0.698	0.727
	test	0.699	0.737

Table 1: Teacher Forcing Ratio and Results

2.4   Varying beam size

We run the seq2seq model with attention model, while also varying the beam size, and recorded the observations.

Results

The observations are compiled in the table below.

Beam Size	Exact Match	Execution
1	0.625	0.692
10	0.657	0.674
20	0.637	0.669

Table 2: Beam Size and Results

A few observations

At first glance, one notices that the accuracy increases on increasing the beam size from 1, which corresponds to greedy search, to a beam size of 10. This is because selecting the next most probable word does not imply that the whole most probable sequence of the words that are about to follow is selected. Thus, the accuracy increases on increasing the beam size.

Another important thing to notice is that the accuracy goes down on increasing the beam size too much. This is because even though this time the set chosen is a superset of the set chosen in beam size 10, the training process in this case would become harder, and in most cases, the beam size of 10 is already sufficiently large to capture most interesting cases.

What we learn from this analysis is to have a sufficiently large, yet small enough beam size to get a good accuracy.