

University of Zurich

Advanced Techniques of Machine Translation

Assignment 1

Omkar Ingale
24-716-466

Task 1

To train on GPU, I added the “—cuda” flag to train and translate script calls in the shell script. CPU training time averaged to 19 seconds per epoch whereas training on GPU took an average of 3 seconds per epoch. This is seen in the images displayed below:

```
Epoch 000: loss 5.458 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 28.13 | clip 1
Time to complete epoch 000 (training only): 19.63 seconds
Epoch 000: valid_loss 5.36 | num_tokens 27.2 | batch_size 100 | valid_perplexity 214 | BLEU 0.000
Epoch 001: loss 4.959 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 25.54 | clip 1
Time to complete epoch 001 (training only): 19.33 seconds
Epoch 001: valid_loss 5.23 | num_tokens 27.2 | batch_size 100 | valid_perplexity 186 | BLEU 0.072
Epoch 002: loss 4.673 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 23.41 | clip 1
Time to complete epoch 002 (training only): 19.28 seconds
Epoch 002: valid_loss 5.07 | num_tokens 27.2 | batch_size 100 | valid_perplexity 159 | BLEU 0.109
Epoch 003: loss 4.353 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 22.34 | clip 1
Time to complete epoch 003 (training only): 19.40 seconds
Epoch 003: valid_loss 4.92 | num_tokens 27.2 | batch_size 100 | valid_perplexity 137 | BLEU 0.144
Epoch 004: loss 4.040 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 22.7 | clip 1
Time to complete epoch 004 (training only): 19.84 seconds
| Validating Epoch 004: 50%|██████████| 2/4 [01:19<01:19, 39.51s/it]slurmstepd: error: *** JOB 23464066 ON u20-chi1vm0-603 CANCELLED AT 2025-10-03T15:03:38 DUE TO TIME LIMIT ***
*
oingal@u20-login-3:~/data/atmt_2025$ sbatch toy_example.sh
Submitted batch job 23464995
oingal@u20-login-3:~/data/atmt_2025$ cat toy_example.out
```

Image 1: Training on CPU

```
Time to complete epoch 000 (training only): 4.66 seconds
Epoch 000: valid_loss 5.36 | num_tokens 27.2 | batch_size 100 | valid_perplexity 213 | BLEU 0.076
Epoch 001: loss 4.901 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 23.46 | clip 1
Time to complete epoch 001 (training only): 3.06 seconds
Epoch 001: valid_loss 5.21 | num_tokens 27.2 | batch_size 100 | valid_perplexity 183 | BLEU 0.091
Epoch 002: loss 4.587 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 21.92 | clip 1
Time to complete epoch 002 (training only): 3.27 seconds
Epoch 002: valid_loss 5.06 | num_tokens 27.2 | batch_size 100 | valid_perplexity 157 | BLEU 0.153
Epoch 003: loss 4.279 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 22.58 | clip 1
Time to complete epoch 003 (training only): 3.01 seconds
Epoch 003: valid_loss 4.93 | num_tokens 27.2 | batch_size 100 | valid_perplexity 139 | BLEU 0.139
Epoch 004: loss 3.995 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 25.14 | clip 1
Time to complete epoch 004 (training only): 3.05 seconds
Epoch 004: valid_loss 4.84 | num_tokens 27.2 | batch_size 100 | valid_perplexity 127 | BLEU 0.280
Epoch 005: loss 3.759 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 26.94 | clip 1
Time to complete epoch 005 (training only): 3.13 seconds
Epoch 005: valid_loss 4.75 | num_tokens 27.2 | batch_size 100 | valid_perplexity 116 | BLEU 0.321
Epoch 006: loss 3.522 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 26.01 | clip 1
Time to complete epoch 006 (training only): 3.01 seconds
Epoch 006: valid_loss 4.68 | num_tokens 27.2 | batch_size 100 | valid_perplexity 108 | BLEU 0.490
Epoch 007: loss 3.308 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 26.4 | clip 1
Time to complete epoch 007 (training only): 3.37 seconds
Epoch 007: valid_loss 4.63 | num_tokens 27.2 | batch_size 100 | valid_perplexity 102 | BLEU 0.216
Epoch 008: loss 3.114 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 27.59 | clip 0.9688
Time to complete epoch 008 (training only): 3.36 seconds
Epoch 008: valid_loss 4.58 | num_tokens 27.2 | batch_size 100 | valid_perplexity 97.9 | BLEU 0.286
Epoch 009: loss 2.919 | lr 0.0003 | num_tokens 29.49 | batch_size 31.25 | grad_norm 27.93 | clip 0.9688
Time to complete epoch 009 (training only): 2.80 seconds
Epoch 009: valid_loss 4.53 | num_tokens 27.2 | batch_size 100 | valid_perplexity 92.3 | BLEU 0.651
```

Image 2: Training on GPU

Task 2

The output file for this task is included in the submission folder. The graph for this task are as follows:

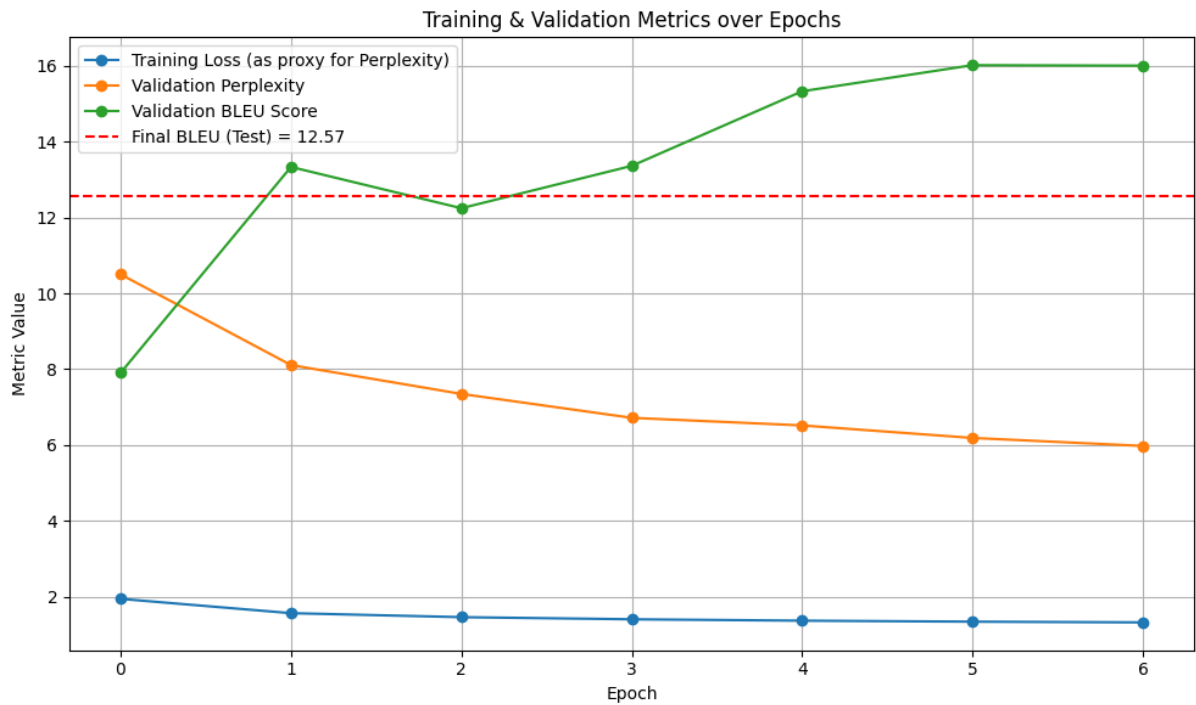


Image 3: Metrics Progression with Epochs (Training Loss shown instead of Training Perplexity)

Task 3

For this task, I chose to use 5000 entries from the WikiMatrix dataset on Opus¹. The final BLEU score on the dataset was 2.05. The output file is included in the submission folder. The way I interpret the result is that the translation model didn't generalize well on the Slovak to English translation task. Most translations are incorrect and word alignment and sentence structure are incorrect.

¹ <https://opus.nlpl.eu/WikiMatrix/sk&en/v1/WikiMatrix>