

CS310 Natural Language Processing - Assignment 3 Recurrent Neural Networks for Language Modeling

Task3

Train the model

Here we train both models with 31105 `batch` (90% size of the whole dataset) and with 1 `epoch` (the main reason is just that my computer lacks a high-performance graphics card, doing one epoch takes about 2 hours for each model).

We use the following hyperparameters for both models:

- `embed_size=128`
- `hidden_size=256`
- `num_layers=2`
- `bidirectional=False`

The loss of RNN:

```
1 2025-03-31 00:59:20,878 - INFO - Starting training with 1 epoch(s), log_interval=100
2 2025-03-31 00:59:25,245 - INFO - Epoch 1, Batch 100/31105, Batch Loss: 8.2583, Avg
  Loss So Far: 4.2274
3 2025-03-31 00:59:29,610 - INFO - Epoch 1, Batch 200/31105, Batch Loss: 8.2716, Avg
  Loss So Far: 4.2483
4 2025-03-31 00:59:34,665 - INFO - Epoch 1, Batch 300/31105, Batch Loss: 7.3097, Avg
  Loss So Far: 4.2720
5 2025-03-31 00:59:40,215 - INFO - Epoch 1, Batch 400/31105, Batch Loss: 8.2487, Avg
  Loss So Far: 4.2776
6 2025-03-31 00:59:45,720 - INFO - Epoch 1, Batch 500/31105, Batch Loss: 7.5610, Avg
  Loss So Far: 4.2888
7 ...
8 2025-03-31 01:27:22,205 - INFO - Epoch 1, Batch 30700/31105, Batch Loss: 3.6060, Avg
  Loss So Far: 3.8673
9 2025-03-31 01:27:27,698 - INFO - Epoch 1, Batch 30800/31105, Batch Loss: 3.6954, Avg
  Loss So Far: 3.8665
10 2025-03-31 01:27:33,157 - INFO - Epoch 1, Batch 30900/31105, Batch Loss: 3.5182, Avg
  Loss So Far: 3.8657
11 2025-03-31 01:27:38,666 - INFO - Epoch 1, Batch 31000/31105, Batch Loss: 3.4214, Avg
  Loss So Far: 3.8649
12 2025-03-31 01:27:44,177 - INFO - Epoch 1, Batch 31100/31105, Batch Loss: 3.5912, Avg
  Loss So Far: 3.8641
13 2025-03-31 01:27:44,460 - INFO - Epoch 1 Summary, Avg Loss: 3.8641
14 2025-03-31 01:27:44,506 - INFO - Model saved to rnn_model_new.pth
15 2025-03-31 01:27:44,507 - INFO - Training completed
```

The loss of LSTM:

```
1 2025-03-31 00:59:31,485 - INFO - Starting training with 1 epoch(s), log_interval=100
2 2025-03-31 00:59:39,997 - INFO - Epoch 1, Batch 100/31105, Batch Loss: 9.9504, Avg
  Loss So Far: 4.8911
3 2025-03-31 00:59:48,342 - INFO - Epoch 1, Batch 200/31105, Batch Loss: 9.9493, Avg
  Loss So Far: 4.9123
4 2025-03-31 00:59:56,626 - INFO - Epoch 1, Batch 300/31105, Batch Loss: 8.9202, Avg
  Loss So Far: 4.9277
5 2025-03-31 01:00:05,031 - INFO - Epoch 1, Batch 400/31105, Batch Loss: 9.0820, Avg
  Loss So Far: 4.9378
6 2025-03-31 01:00:13,432 - INFO - Epoch 1, Batch 500/31105, Batch Loss: 8.3582, Avg
  Loss So Far: 4.9457
7 ...
8 2025-03-31 01:37:20,846 - INFO - Epoch 1, Batch 30700/31105, Batch Loss: 3.6006, Avg
  Loss So Far: 4.1351
9 2025-03-31 01:37:26,392 - INFO - Epoch 1, Batch 30800/31105, Batch Loss: 3.4709, Avg
  Loss So Far: 4.1333
10 2025-03-31 01:37:31,946 - INFO - Epoch 1, Batch 30900/31105, Batch Loss: 3.6718, Avg
  Loss So Far: 4.1315
11 2025-03-31 01:37:37,480 - INFO - Epoch 1, Batch 31000/31105, Batch Loss: 3.4123, Avg
  Loss So Far: 4.1297
12 2025-03-31 01:37:43,010 - INFO - Epoch 1, Batch 31100/31105, Batch Loss: 3.5172, Avg
  Loss So Far: 4.1280
13 2025-03-31 01:37:43,276 - INFO - Epoch 1 Summary, Avg Loss: 4.1280
14 2025-03-31 01:37:43,324 - INFO - Model saved to lstm_model_new.pth
15 2025-03-31 01:37:43,324 - INFO - Training completed
```

The final loss of both models is about **3.5**. It's rather average but sufficient.

Perplexity scores on the test set

Here is the comparison of the perplexity of the two models(during the test process):

Total non-padding tokens: 1105920	Total non-padding tokens: 1105920
Total loss: 3930417.2094	Total loss: 3894121.2832
Average loss: 3.5540	Average loss: 3.5212
Perplexity: 34.95	Perplexity: 33.82
RNN Test Perplexity: 34.95	LSTM Test Perplexity: 33.82

Their perplexity and loss are quite close.

Generate 5 pairs of sentences using greedy search

RNN:

```
1 Generating 5 pairs of sentences (RNN):
2
3 Prefix: Harry looked
```

4 RNN: harry looked up at the ceiling and then the door swung open and he was sure that
he had not seen him .he was wearing a long overcoat and a

5

6 Prefix: Hermione said

7 RNN: hermione said ron nodding toward the remainder of the bus .the ministry of magic
confirmed that he had been discovered to act as though he had a stitch in his

8

9 Prefix: Ron shouted

10 RNN: ron shouted at the dangling and carrying a large suitcase and banged her eyes
and scanning it to the kilted glass of the chamber of secrets and finally the dark

11

12 Prefix: Dumbledore stood

13 RNN: dumbledore stood up and croaked hedwig and roger davies was almost glad to see
her .aberforths not supposed to be in the forest .harry felt a thrill of foreboding
.avada

14

15 Prefix: Snape glared

16 RNN: snape glared at him as though he had a stitch in his chest was torn and his eyes
were rolling madly and down the table and shouted expelliarmusv and he

1 Generating 5 pairs of sentences (LSTM):

2

3 Prefix: Harry looked

4 LSTM: harry looked at him and saw a sliver of silverwhite shining brightly as she
threw herself out of the room by the looks of her nose was gone.crack.james

5

6 Prefix: Hermione said

7 LSTM: hermione said ron nodding at her anxiously and harry could tell that she was a
very handsome woman she was clutching his wand in his pockets.he had no idea

8

9 Prefix: Ron shouted

10 LSTM: ron shouted at him as though she was steeling herself and i think so said
hermione in a low voice and harry was pleased to see that she was a

11

12 Prefix: Dumbledore stood

13 LSTM: dumbledore stood there with her wand pointing at her earlobes.i suppose he can
say i was going to be a bit more trustin said ron.i think moms got

14

15 Prefix: Snape glared

16 LSTM: snape glared at him and harry and hermione both looked at harry and said
nothing of her flinching and in case he could never talk about the dursleys but he

We can compare the two results in the following dimensions:

- **Coherence:**

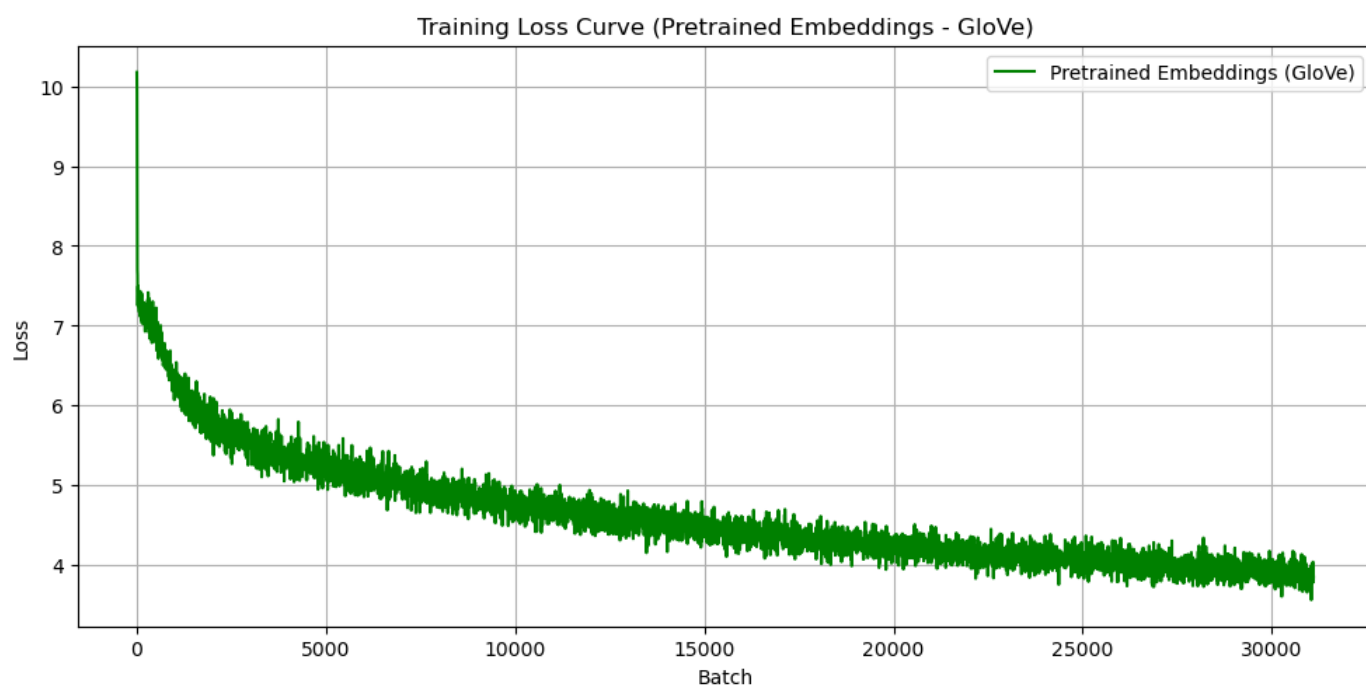
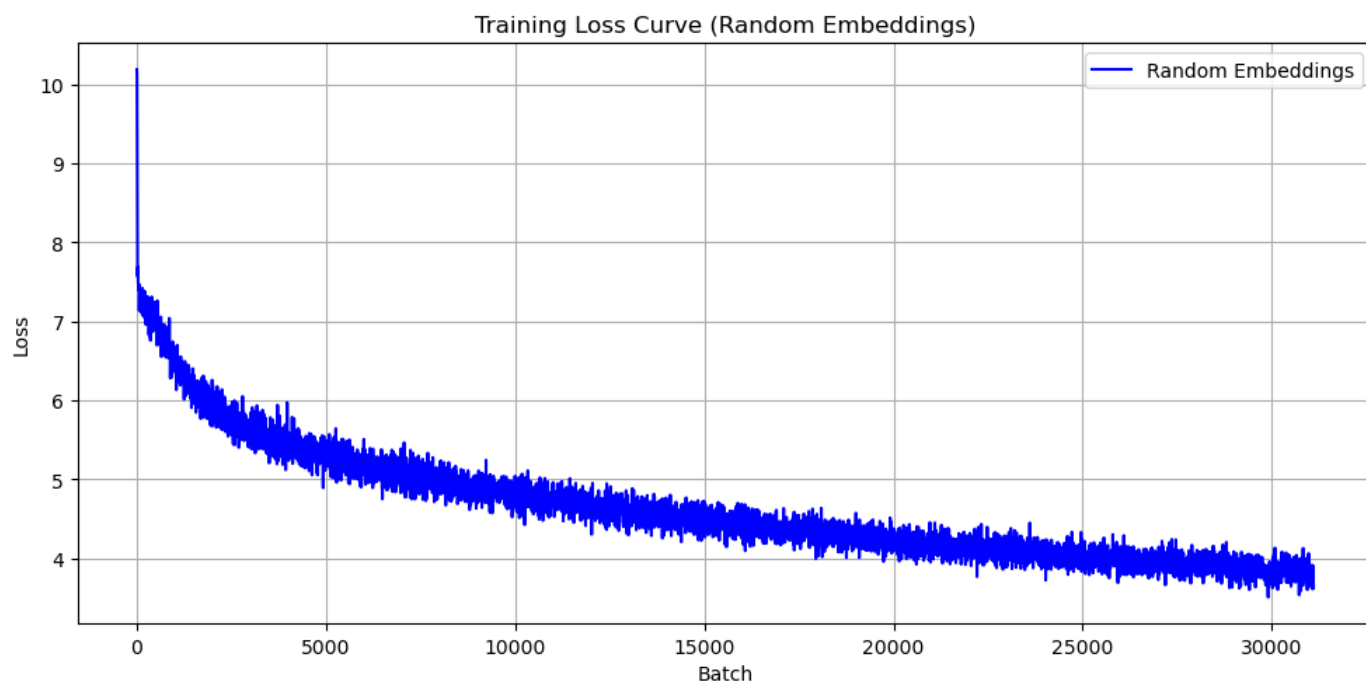
- LSTM starts strong and keeps things clear in the beginning and middle, but it gets messy at the end of longer sentences, like with "Dumbledore stood," where it doesn't make sense anymore.
- RNN begins okay but quickly gets confusing and jumps around, like in "Ron shouted," where the ideas don't connect.

- **Grammatical Correctness:**
 - LSTM makes sentences that mostly sound right, but it often mixes up "he" and "she" or "his" and "her," which is weird.
 - RNN's sentences are messier and harder to follow, but it doesn't mix up pronouns as much.
- **Thematic Relevance:**
 - LSTM stays closer to the "Harry Potter" story, using characters and magic stuff like "wand."
 - RNN drifts off and throws in random words like "kilted glass" that don't fit the story at all.
- **Character Consistency:**
 - LSTM shows more about characters and how they talk or act, but it gets confused about whether they're male or female.
 - RNN keeps characters simpler but often makes them do or say things that don't make sense.
- **Overall Quality:**
 - LSTM makes longer sentences that feel more like "Harry Potter," but sometimes they don't finish well.
 - RNN's sentences are shorter, more random, and not as good. LSTM is better, but it still needs some work on training and the data to fix the problems.

Task4

The training loss curves

We pick up the loss every 500 batch, otherwise there will be too many dots and hard to distinguish how much is the loss. The loss during the training process is quite similar.



Final perplexity scores on test set

Final Perplexity (Random Embeddings): 42.45

Final Perplexity (Pretrained Embeddings): 45.56

The final perplexity of the LSTM model with pretrained embeddings is a bit larger than that with random embedding, which is surprising.