

NLP Assignment 3

20191138 Hyungyu Lee

Problem 1

I implemented LSTM following the given restrictions. The code passed the assertions and shell script test.

Problem 2

Choosing Training Dataset

Min Hee Jin's Press Conference (Discarded)

On April 25th, 2024, Min Hee Jin, CEO of ADOR and producer of famous KPOP group, NewJeans, had a press conference regarding her conflict with ADOR's parent company, HYBE. HYBE raised an allegation that Min, who holds 18% of ADOR's share, tried to overthrow HYBE's governance towards the company, and tried to break contracts between ADOR(HYBE) and NewJeans. HYBE demanded board of directors be held so that they(who holds 80% share of ADOR) can fire Min. In the two-hour-long unprecedentedly unconventional conference, Min was really honest, emotional and used lots of slurs and the conference became a meme. I wanted to create sentences from '미안하지만 (I am sorry but)' in relation to the part of the press conference that became most popular, "아저씨들이, 미안하지만 🧓저씨들이 (Those boomers, I am sorry but, those old creepy boomers)". There was a transcript of the whole press conference on the internet, and I cross checked and used that .txt file to train the model. But the model generated texts like these:

{ 'generated_text': '미안하지만 0000 00 000000 000 0 00 00 000 00 0 00 0 000000 0 0000 00 000 0 00000 0
{ 'generated_text': '미안하지만 0 0 00 0 0000000000 00 000000 00 00000000 00 000000 00000000 00 000 0 00
{ 'generated_text': '미안하지만 0 00 0 닳 000 000 0000 0000000000 000 000 0000 0 00000 0 000000 00000 000
{ 'generated_text': '미안하지만 0 00 0 000 000 000 0 000 000000000000000 닳 0 닳 0000000 000 000 00000 000000
{ 'generated_text': '미안하지만 00 000 0 0 0 00 00 0 0 000 000000 00 000000000000 00 000000 000 00000000 0
{ 'generated_text': '미안하지만 000 0 000000 0 0000 000 00 00 000 00 0000 0 000 000000 00000000000 닳 0000
{ 'generated_text': '미안하지만 0000담 0익 000 0 00 0 0000000 000 0000 00 000 0 00000 00000 00000 0000 0000
{ 'generated_text': '미안하지만 0 000 0 0 0 0000 000 00000000 0000 00 0000 00 00000000000 0 000
{ 'generated_text': '미안하지만 0 000 0 0000000 000 00 00 0 0해 0 00 00 00 0000 0000000 00 000 000 00
{ 'generated_text': '미안하지만 0 000 00000 00 000 000 0000000 닳 0 00 00 00000000 0 00000 000000 00 000

Thus, I could not use this corpus. I was thinking of translating the corpus to English, generate the text, and translating the generated text to Korean again, but that would not be an emotional and expressive text like the press conference, and doing that process itself will be equivalent to creating another model. So I had to discard this corpus.

Novel *Twilight* by Stephanie Meyer

Another corpus that I had in mind was the novel *Twilight* by Stephanie Meyer. *Twilight* is a young adult fantasy novel about Bella Swan, a teenager who falls in love with Edward Cullen, a vampire. The book, which is the first in the *Twilight Saga*, became famous for its intense romance and notoriously cringey moments. These include Edward watching Bella sleep without her knowledge, and overly dramatic dialogue such as "You are my life now." Despite—or perhaps because of—its melodramatic and sometimes awkward elements, *Twilight* captivated a massive audience and sparked a popular film series. The actors who played Edward Cullen and Bella Swan, Robert Pattinson, and Kristen Stewart, both made many critical remarks regarding the book. Having read the whole series and watched the movies while growing up, I wanted to see if the model can be fine-tuned to generate the cringey parts like below:

"You are an idiot," he agreed with a laugh. Our eyes met, and I laughed, too. We laughed together at the idiocy and sheer impossibility of such a moment.

"And so the lion fell in love with the lamb ..." he murmured. I looked away, hiding my eyes as I threw the word.

"What a stupid lamb," I sighed.

"What a sick, masochistic lion." He stared into the shadowy forest for a long moment, and I wondered where his thoughts had taken him.

I got the corpus in plain text format from [Internet Archive](#), deleted parts that are not the novel itself (table of contents, ISBN, etc.).

Difficulties

```
OutOfMemoryError: CUDA out of memory. Tried to allocate 96.00 MiB. GPU 0 has a total capacity of 14.75 GiB of which 25.06 MiB is free. Process 12257 has 14.72 GiB memory in use. Of the allocated memory 14.50 GiB is allocated by PyTorch, and 90.10 MiB is reserved by PyTorch but unallocated. If reserved but unallocated memory is large try setting PYTORCH_CUDA_ALLOC_CONF=expandable_segments:True to avoid fragmentation. See documentation for Memory Management
```

Because the corpus had about 670,000 letters and 120,000 words, Google Colab had several `OutOfMemoryError`. I attempted to fix it first with code that frees memory:

```
torch.cuda.empty_cache()
import gc
gc.collect()
```

The code seemed to work fine, but had the same error after several iterations. I had to decrease the `batch_size` to 24 to avoid error.

Code to Train the Model

```
from tqdm import tqdm

torch.cuda.empty_cache()
import gc

optimizer = torch.optim.Adam(model.parameters(), lr=1e-2)
loss_record = []
model.cuda()
num_iteration = 500 # you can change num updates here

train_iter = iter(train_loader)

for i in tqdm(range(num_iteration)):
    try:
        batch = next(train_iter)
    except StopIteration:
        train_iter = iter(train_loader)
        batch = next(train_iter)
    out = model(batch.cuda()).logits
    out = torch.log_softmax(out, dim=-1)
    correct_next_word = batch[:, 1:]
    out_flatten = out[:, :-1, :].reshape(-1, out.shape[-1])
    correct_next_word = correct_next_word.reshape(-1)

    log_prob_of_correct_word = out_flatten[torch.arange(len(correct_next_word)),
correct_next_word]

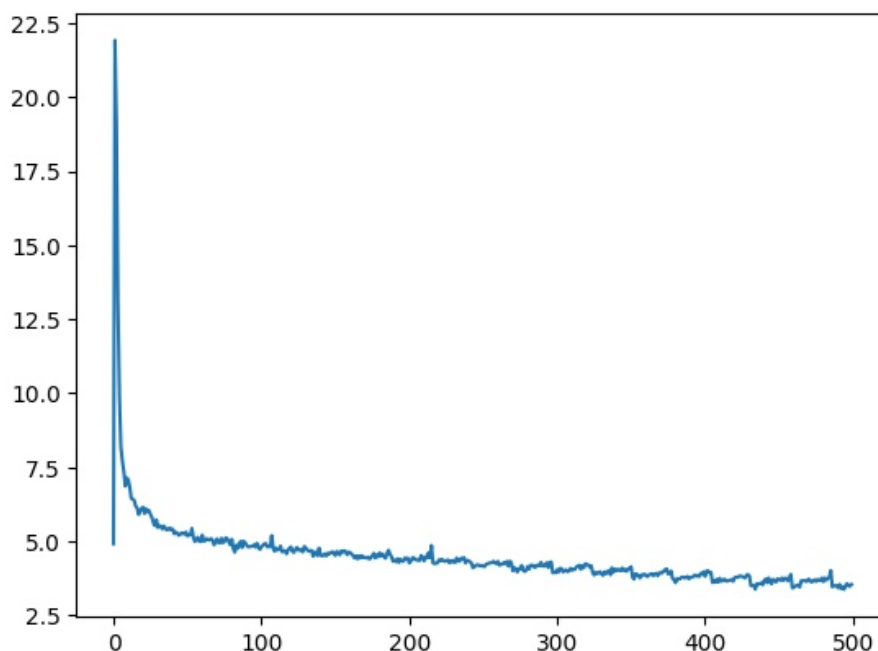
    loss = -log_prob_of_correct_word.mean()
    loss.backward()

    optimizer.step()
    optimizer.zero_grad()
    loss_record.append(loss.item())

plt.plot(loss_record)
```

- Adam `optimizer` is called, with `model`'s parameters, and learning rate of 1e-2.
- Then, `model.cuda()` moves the model data to GPU for accelerated computing.
- `train_iter = iter(train_loader)` initializes iterator over training dataset.
- For 500 iterations, try and except structure detects whether the `dataloader` runs out of data to work with (`StopIteration`). If so, we reinitialize the iterator.
- `batch.cuda()` moves the data to GPU, and then to the `model`, and gets the output. The raw output is processed through softmax.
- The target correct next words for each sequence in the batch are isolated with `batch[:, 1:]`.
- `out_flatten` reshapes the output tensor to align with these targets, flattening it to match the batch's sequence of labels.

- `log_prob_of_correct_word` gathers the log probabilities corresponding to the actual next words in the sequences.
- `loss = -log_prob_of_correct_word.mean()` computes the negative mean of these log probabilities, which corresponds to the Cross-Entropy loss, a common loss function for classification tasks like language modeling.
- Then, we do backpropagation.
 - `loss.backward()` computes the gradient of the loss with respect to the model parameters.
 - `optimizer.step()` updates the model parameters based on the gradients.
 - `optimizer.zero_grad()` resets the gradients of the model parameters to zero for the next iteration.
- The loss value for each iteration is appended to `loss_record` for later analysis.
- After training, the losses recorded during training are plotted using `matplotlib` to visualize the training progression. The loss looked like below. I think doing about 400 iterations will be enough.



Interesting Examples

Since the original text is known to be really cringey and teenage-like, I anticipated some weird usage of words, which is inevitable when using language model, to be seen as a concept of the text. Thus, I neglected some syntax errors. I used `your_max_length = 30`, and included some interesting examples in the report.

Prompt: `I am not` - General Prompt

Some created texts were random gibberish, whereas some were interesting and cringey as I intended. `I am (not) a little little` was the most frequent pattern.

- "I am not really know." He raised to his hands released his lip, trying to the back of it with my way, and then.
 - This text captures a moment of vulnerability and hesitation, which is engaging in narrative fiction.
 - But, sentence structure is awkward, which shows the model struggling with verb phrases and prepositions. This compromises grammatical correctness and clarity of the sentence.
- "I am not going by. It is that we smells like the same and that she called my waist in place in his arms there's no possibility"
 - This example includes sensory details ("smells like the same") and hints at intimacy ("called my waist in place in his arms"), which enrich the narrative.
 - But still, the model suffers from grammatical errors and confusing constructions ("smells like the same and that she called my waist in place"). This illustrates the model's need for improvements in syntax and pronoun agreement to enhance coherence.
- "I am not attentive as I've never have an assigned bathroom. "Cullens, falling resgh has to order are all top of intense"
 - This text shows creative use of context and character-specific references, though it is significantly hindered by nonsensical phrasing ("falling resgh has to order are all top of intense"). The sentence structure is wrangled, reflecting issues with logical progression and word choice, which hinders overall narrative comprehensibility.

Prompt: I am not a vampire. You are - Themed Prompt

- 'I am not a vampire. You are you.' "My mother's am about that I was too late to tell Charlie, about knowing Charlie wasn't. I have to bring to sit with my'
 - This text effectively conveys emotional urgency and hints at complex relationships. The model successfully generates a cliffhanger, making the narrative intriguing. The beginning of this sentence starts strong with a clear emotional conflict.
 - However, the phrase "My mother's am about that" shows a grammatical misalignment. The model struggles with possessive constructions ("My mother's") followed by verb forms, leading to a confusing and grammatically incorrect output.
- 'I am not a vampire. You are you, or I guess it's diabolical about your thoughts.' He put his breath he was an unpleasant points me, with details. "Just really'
 - This example shows an ambitious attempt by the model to handle complex dialogue and emotional description. However, the phrase "He put his breath he was an unpleasant points me," although evocative, is syntactically incorrect. The model fails to maintain grammatical structure, leading to a breakdown in coherence.

Prompt: Bella, I love - Main Character Prompt

I anticipated that sentence with Bella will be well constructed - both in terms of meaning and grammar, since Bella is the main character and mentioned many time in the corpus. But many of them were randomized, leading to the point where the meaning of sentence is rarely preserved.

- 'Bella, I love me recklessly. And then Mike promised. "He has weapons for that." His voice was leaning for the distance. "Something." "As yourself," Dr. "'

- It can be said that the model creatively mixes direct speech with narrative elements, showing potential for dramatic and emotionally resonant storytelling. But, it also illustrates typical issues such as abrupt changes in subject and unclear references ("His voice was leaning for the distance"), suggesting that the model may struggle with maintaining consistent narrative perspectives and contextually appropriate expressions.

Prompt: Bella and Edward went to - Main Character Prompt

In the book, Bella and Edward go to see a baseball game. This is not just any ordinary baseball game, however. The Cullen family plays vampire baseball during a thunderstorm, as the sound of the ball being hit is so loud that it resembles thunder. This scene is significant as it leads to a crucial plot development where Bella is exposed to other vampires who are not as restrained as the Cullens, posing a threat to her safety. This event escalates the tension and the danger in the story, leading to further dramatic developments.

- 'Bella and Edward went to find an angel, Jasper's high, Jess corrected in response. And someone's still...' I sighed onto my eyes carefully, around my throat. Do you promised'
 - This text is worth noting because, in terms of narrative, it weaves in multiple characters from the "Twilight" universe, enhancing the sense of a connected narrative. Jasper and Jess are mentioned, which grounds the text in the series' character dynamics. The phrase "sighed onto my eyes carefully, around my throat" evokes a sense of intimacy and potential vulnerability, fitting for the heightened emotional stakes typically found in "Twilight." But still, grammatically, this sentence is hard to understand.
 - It can be said that this model is failing in terms of writing *good* sentences, or succeeding in writing in *creative* themes.

Prompt: At Starbucks, Bella - Unknown Background Prompt

Starbucks do not appear in the book. So, I wanted to see what happens with the prompt, and observe how the model deals with unknown background. The generated texts were more complex and hard to understand than other prompts.

- 'At Starbucks, Bella," I resented office, like me as far from the rest of First, the sound of me, and then you can't laugh, watching the rest of her, when he'
 - This text combines a mixture of emotions and a somewhat unconnected narrative, like other generated texts. The abrupt end raises questions and could act as a cliffhanger in a larger narrative context. If this sentence was in the *Twilight* book, it would be considered as really strangled sentence.
- 'At Starbucks, Bella. Don't saved you're here or his name. "Do you your mom." "Why what he's just with Mom," my thoughts." "I told."'
 - The phrasing and grammar are somewhat jumbled, reflecting the spontaneity or stress of the conversation, which, while confusing, adds a realistic touch to how people might speak under stress.

Prompt: At cafeteria, Bella - Known Background Prompt

Cafeteria exists as a background in the novel. Thus, I wanted to see if the content or event is influenced with the original plot of the novel. But that tendency did not exist.

- "At cafeteria, Bella. He was taking me, he took his perfectly still worried about him. He was very long time, smiling, and he was free to the edge of his, if he was"
- "At cafeteria, Bella. When I was still flaring, and "So, but it for me like it." He frowned at it. "What?" "It's go?" he said"

Prompt: Edward needed to drink blood. He went to the school and - Main Character Action Prompt

I wanted to see if Edward will go drink blood, which is easily derived action from the prompt, and also connected to the novel's main theme(vampire). But, Edward actually does not drink blood from human (the Cullens refer themselves as *vegan* for this.) Regardless, I wanted to see if drinking or consuming blood appears in the generated text. Initially, I set `your_max_length = 40`, but no such action appeared. I then increased `your_max_length` to 100 and tried again.

- 'Edward needed to drink blood. He went to the school and the other with him. He waited for his unusual beneath his reaction or one of it on his own mind, probing for my emotions.'

Still, there was no example where Edward drinks blood. The act of drinking blood, especially from humans, does not occur in the novel, and blood itself is rarely discussed in the plot. Instead, aspects like eternal life, diamond-like skin, and incredible strength are emphasized. I had hoped the model might learn from these rare examples or at least connect to the vampire theme, but it did not work as intended.

Insights

After experimenting with numerous prompts, I could see both the strengths and limitations of the model. The model typically generates text that aligns with specific characters and settings, or reflects the style of *Twilight*. It effectively portrays complex emotional states and interpersonal dynamics, enriching storytelling. Although this may be created from noise or randomness, such complex structure is something that cannot be easily created by human authors easily.

However, the model exhibits lots limitation as well. The model struggled with maintaining coherence and grammatical accuracy, especially in complex sentences. This affected the readability and immersion of the narrative. Also, while the model can initiate and continue conversations, it occasionally fails to maintain consistent character voices, potentially disrupting the narrative flow and diminishing character authenticity. But again, this is also one feature of *Twilight* corpus.

Also, when faced with unfamiliar settings not explicitly covered in the training data, Starbucks, for example, the model's performance can be uneven, impacting the depth and authenticity of the generated content in these new contexts.

I wondered if the corpus size was a contributing factor. I downloaded all 4 books in *Twilight* Saga (*Twilight*, *New Moon*, *Eclipse*, and *Breaking Dawn*), and did same training. However, overall generated texts were mere similar, both in terms of narrative and syntax. The dizzy, tangled and highly expressive language tone of the saga may be the contributing factor influencing such generation.

Can the Model Discern a/an?

Prompt: You are a

```
{'generated_text': 'You are a promise to get a sidewalk," I rolled his eyes and sighed. his eyes, '},  
{'generated_text': 'You are a promise to come." I held his truck drop it was. "Oh'},  
{'generated_text': 'You are a monster?" Mike asked, his gleaming his few more to his lap, and I'},  
{'generated_text': 'You are a monster?" I wonder his question. "Enough'},  
{'generated_text': 'You are a promise." I leaned my door, and his eyes appraising me again as usual'},  
{'generated_text': 'You are a promise." Mike peeked back to the couch, his voice in his voice "},  
{'generated_text': 'You are a promise to carry you stories." I rolled his eyes deep breath. It was a needle'},  
{'generated_text': 'You are a promise," he announced, his voice, his hps pressed my hands. '},  
{'generated_text': 'You are a promise to get it," I said. I guess. His voice was still'},  
{'generated_text': 'You are a promise to come." I watched as he spun.
```

Prompt: You are an

```
{'generated_text': 'You are an misty," I asked my five. He grinned. "It's still,"},  
{'generated_text': 'You are an misty." accelerating, and I was a foot of his her car, my '},  
{'generated_text': 'You are an cold — me." "I'm right?" those." He drop he'},  
{'generated_text': 'You are an misty." I thought I really," he couldowled. '},  
{'generated_text': 'You are an misty about his brighter twenty-are pattern of the van that I gunned my scent'},  
{'generated_text': 'You are an misty about his chance." I was a drop of his attention to see his own'},  
{'generated_text': 'You are an angry, still brighter pieces; I was going to find so that I will be with'},  
{'generated_text': 'You are an misty." "Oh," I said. "I really think'},  
{'generated_text': 'You are an misty." I said as we were still stealing my He smiled.'},  
{'generated_text': 'You are an understatement." I frowned.. as I sighed. "I said my'}}
```

I conducted an experiment to determine whether the model could accurately discern and appropriately use words that start with consonants versus vowels. Specifically, I tested the phrases 'You are a' and 'You are an' to see if the model would correctly follow them with words beginning with consonants and vowels, respectively. The results were inconclusive: while 'a' was often followed by words starting with consonants, 'an' was inconsistently followed by both consonant- and vowel-starting words.