Sarah Groark
Generative AI: Assignment 5
LSTM

LSTM Training on Jane Austen Text

The assignment called for training of an LSTM on three different Jane Austen works in hopes of generating text in her style of writing. The three books included *Pride and Prejudice, Emma,* and *Sense and Sensibility.* Due to the sensitivity of an LSTM to varying properties, a total of four models were trained with different parameters. Each of these models will be explored here:

Model 1 =

Model 1 featured the most basic setup of the four – one LSTM layer with 128 N_UNITS trained over the course of 25 epochs. The model summary reported 3.1 million parameters to be trained on in the model.

After the conclusion of training, the following prompts were given to the model at various temperatures:

1. Temperature = 0.25 – "Emma saw that"
   a. Generated text = `Emma saw that he had been in town , and was the time of her own , and the most natural and frozen maid .`
2. Temperature = 0.5 – "Pride had taken it all"
   a. Generated text = `Pride had taken it all over in her power to be done .`
3. Temperature = 1.0 – "How is it that"
   a. Generated text = `How is it that i have not occurred to him ?`

It can be seen that as the temperature value increases, the coherence of the sentence increases as well. The generated sentence "How is it that i have not occurred to him ?" is the most sensible and stylistically accurate of the three. The other two utilize word choices that don't make much logical sense to a reader. However, it can be seen in the first clause of the first sentence that the generator strung together a plausible clause that makes sense (Emma saw that he had been in town), however, as the sentence continues, the coherence declines. It should be noted that any sentence generated by the models here may not explicitly depict Austen's style of writing, as the models used to generate them were trained for a short amount of time.

Model 2 =

Model 2 featured two LSTM layers, instead of just one, but with the same amount of N_UNITS as seen in the first model. This experiment was designed to compare and contrast the coherence and grammatical accuracy of the two models to determine whether the presence of more LSTM

layers would impact sentence quality. This model featured just over 3.2 million parameters, slightly more than the first model, which would be trained on over the course of 25 epochs (same as model #1).

Prompts and generated text:

1. Temperature = 1.0 – "In a situation that"
   a. Generated text = `In a situation that painful solemnity could be in beauty .`
2. Temperature = 0.5 – "In a situation that"
   a. Generated text = `In a situation that is so much preferable to his own .`
3. Temperature = 0.2 – "Jane had given"
   a. Generated text = `Jane had given her pleasure in her own mind , she was not in the least to say that she was not to be supposed`
4. Temperature = 1.0 – "Mr. Darcy wanted"
   a. Generated text = `Mr. Darcy wanted to render their excuse , that my mind left no further amends to own the distress of the feelings and while he`

The presence of just one more LSTM layer had an impact. The cohesiveness of a sentence generated at just 0.5 temperature is more logical than the sentence generated at 0.5 in the model with just one LSTM layer. Similarly, this model's generation to the prompt "In a situation that" at a temperature of 1.0 highlights Austen's writing tendencies (more fluidity and unique vocabulary) more so than model #1. Finally, even the generated sentence at the lowest temperature – which should be the most incoherent – still possessed some qualities of logic (especially in the first clause - 'Jane had given her pleasure in her own mind'). Overall, the presence of even just one more LSTM layer proves that the inclusion of these layers positively impacts the quality of the generated sentences, as the input text is more closely examined in terms of their semantic relationships.

Model 3 =

This model was used to examine the relationship between quality of outputs and the number of units in the LSTM layer (ie. N_UNITS). The N_UNITS represent the hidden state size of the layer and are used to capture complex patterns in the text. The idea remains that the higher quantity of N_UNITS results in more accurate sentences. Additionally, two phrases were each prompted into the generator each at two different temperatures (0.5 and 1.0), to analyze the impact of temperature with the higher number of units.

Prompts and generated text:

1. Temperature = 1.0 – "I was certain that"
   a. Generated text = `I was certain that it was impossible for sir john there .`
2. Temperature = 0.5 – "I was certain that"
   a. Generated text = `I was certain that it was not so .`
3. Temperature = 1.0 – "Emma said that"
   a. Generated text = `Emma said that he was not in their conclusions .`
4. Temperature = 0.5 – "Emma said that"
   a. Generated text = `Emma said that he would not allow him to be the best of the gentleman .`

This model resulted in a slightly higher level of grammatical awareness than did model #1 (same amount of layers, but lower N_UNITS). The use of nouns after verbs is a bit more consistent, as well as the subject placement in the sentence itself (ie. consistent placing of 'it' or 'he' or other nouns following verbs and articles). Even at lower temperatures, the generated outputs still display some level of coherence, despite only being trained through one layer. Overall, the inclusion of more N_UNITS seems to have aided in the word placement and semantic quality of the generated sentences.

Model 4 =

The final model is a cumulative approach with the goal of taking advantage of the positive impacts of both increased LSTM layers and N_UNITS. This model was trained through two LSTM layers (rather than just one) as well as 256 N_UNITS (compared to just 128).

Prompts and generated text:

1. Temperature = 1.0 – "Mr. Bingley was good-looking"
   a. Generated text = `Mr. Bingley was good-looking enough ; it was not mistaken so ; and did not give off the opportunity of saying to miss bennet , who considered the impression of some others , but by its uncertain situation , that they did not enjoy them with each melancholy .`
2. Temperature = 1.0 – "He danced all the dances"
   a. Generated text = `He danced every dance towards the table , when sir john went on , which must be just supported , and by supplying the sound of the meeting , were all to resist harriet's happiness .`

Evidently, the combination of more LSTM layers and N_UNITS results in lengthier sentences. In a practical setting, this would be a good step towards a more realistic implementation of the model. However, the model was still only trained for 25 epochs, so there are still limitations to the quality of the outputs. Nevertheless, these outputs string together some coherent, plausible phrases, for example – 'And by supplying the sound of the meeting' and 'He danced every dance towards the table'. These two phrases demonstrate sensical awareness – specifically spatial awareness of where Sir John was dancing towards, as well as auditory cognizance, as seen in how the use of 'supplying the sound of the meeting' could make logical sense to a reader in a certain context.

Conclusion

Overall, this is a very small glance into what LSTMs could accomplish. The ability of these four small, limited models to generate bits and pieces of sentences that are coherent and, at times, reflect Jane Austen's writing style, is quite fascinating. To further expand on these models, the implementation of not only significantly more LSTM layers and N_UNITS, but training epochs, would result in seamless, quality sentences in Jane Austen's style. This improvement was seen even in this assignment's small implementation, where the inclusion of just one more LSTM layer and slightly more N_UNITS resulted in a tangible difference.