CharRNN Training

Shakespeare text

Samples using MEDIUM temperature = 0.5 (default)

- **Not using top_p/top_k**

Whim that sollied the mest and man a sold made mise this of the me wit in the man the man the sard the man the man are of you say shall my the and man the mame this om and this and le emame the sard of

- **Using top_p = 0.3**

3 stand in that never the senter.

CASSTO:

I we than sond neve all speath scall now me a come supers, mare that that that netcand this sckester son mand and, fand the actime a that store the a stent! s

- **Using top_p = 0.9**

ZASS:

Hearsine that in that the have have and will the for sonce.

How me a bestent.

CORTO:

Hear tellow will that have and and are to a cand fort well there seave and that that the sistere new men, ca

- **Using top_k = 40 (default)**

A:

I have and the cait.

Hood and as that make and the mone the now the now were,

Of the to the a comes that that now the seart this the that that the me prat a the sise you well sell and in the can to

- **Using top_k = 100**

STRO:

This now, and that the are with and and cance,

I have menter the scall to hone that to that marter to that hence,

CARIO:

Hemence, and for you, that the ston me with to hone no some for to you na

Samples using LOW temperature = 0.1

Y:

I have the have the cand the have that the cand the cand the seep that the the have the seep the cand the senter the cand the seep the seep the seep the seep the cand the seep the seep the cand the

Samples using HIGH temperature = 0.9

While rathery.

DORO:

He, to turtants, a that noblencerblethale, fistitity; tran you hettet: than sune tr he to sey, this

I and Heavavand thil

Ot fether. wisstis wotindst, We niscalt. Hetcones:

To thra

Sherlock Text

Samples using MEDIUM temperature = 0.5 (default)

- **Not using top_p/top_k**

15¶»9Ã9½[999Z9Ã`±99989ª999¶`99999999©999Ÿ9Z999¨99899899©9½99999999999999899 99`9¨9½)999©9999999999±9Ã998ª899¼999´999©9»9)9999Ÿ999®99999±9999999999999)99999999999999`999999â999)9¨899¨98ª99999999999¶9999

- **Using top_p = 0.3**

Le the sher the in the react the glair, I bust is in the she hou surring a the is you I the I dear not the se rrach this

all the all I dear that in she he a the pleak I the reams shave have she the se

- **Using top_p = 0.9**

I dou have bremark I all I dee shat is not the the don't is I the manting the reetill that that man's a the rear man's sextime that I deet be all the see the have the sex," It in the rear it been."

- **Using top_k = 40 (default)**

My exce is I a as up ar the the sear the may."

"The kngerm all is I and that I the she in the the the be not is not the rearked you rears, but the the brisher the the up the blan you breat it's as bu

- **Using top_k = 100**

I all the stain that that is side is I the slay as near the have the manded youm be in sear. It the but that the man ther the the but the blears and the dear the splen have brist the may be all hear

Samples using LOW temperature = 0.1

"The she rear the she she she she she sear in the she she she the she she she she the the she she she she sear the she she she rear the she but the she the rear the she she she the she she she she s

Samples using HIGH temperature = 0.9

uph, gividusias is I are tole beet I all goin slead buts."

"I ad all is to jhe in sherre, the seethouble bood am I her thor I showin--wirain. "The segh. Drithem in youse," It's ain I she unce," brems

CharLSTM Training

Shakespeare text

NIO:

Nay, and be thee some a pring my shou mis a have that here of a be pore in that you shou a the son a feellister a be in a deat, I as thou ablive a be lain frieng shou she a bair and she should my

Semen her hall the man but me a bracke with the by bet the the man.

ANTONY:

Sherlock text

Ã:

"Yound the fack me for a more his a will had be a chere away I conger."

"We colled the prom which a last the wore will be at dear and the brow and his me her and a pars, some your his the a look

Comparison between RNN and LSTM

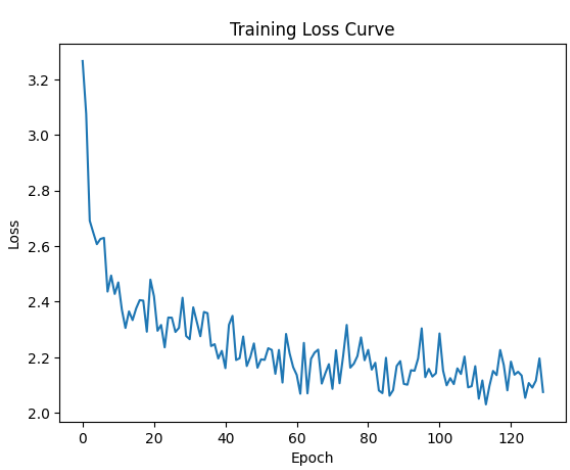

Figure 1:CharRNN Loss

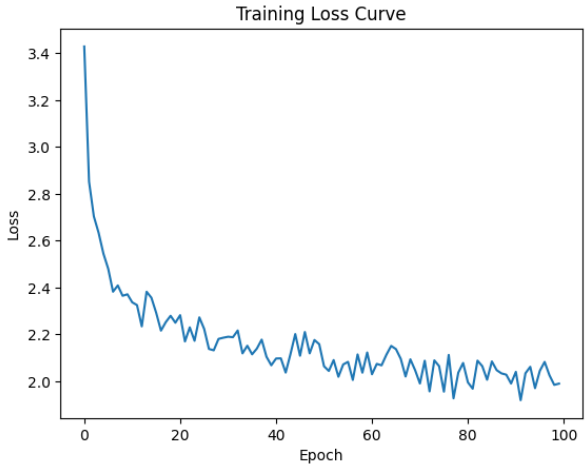

Figure 2: CharLSTM Loss

Training-wise, we can notice that LSTM has a lower loss and more weighted curve in the decreasing direction. RNN has a lot of variance and its loss ping-pongs more than LSTM. In terms of speed, I noticed LSTM was a bit faster than RNN but not by much when trained with 130 epochs and 10 examples per epoch.

Coherence-wise, LSTM seems to produce more coherent and grammatically correct sequences compared to RNN. The generated text from LSTM generally follows a logical structure with recognizable words and phrases. On the other hand, RNN produces text that is less coherent and sometimes contains nonsensical words or phrases, even symbols and numbers.

In terms of understanding and context, LSTM looks to have a better understanding of context. It can maintain context within sentences and generate text that is more consistent with the style and language of the input text, although it is quite broken. RNN fails more frequently in keeping consistent context and this results in more disjointed and random sequences.

In both RNN and LSTM, the generated text from the Shakespeare dataset generally maintains the style and tone of Shakespearean language, with archaic words and poetic expressions,

When talking about the Sherlock text, while attempting to mimic the style of Sherlock Holmes narratives, seems less consistent. LSTM generates sequences that capture some essence of Sherlock Holmes' language and tone, but RNN struggles and is more chaotic, easily observed by the multiple texts that mostly contain the digit ‘9’.

CharRNN Results Evaluation

Temperature

Low temperature: Very deterministic text generation, we can see many repeated sub-sequences in the samples. There's minimal variation in word choice, and the generated text closely resembles the training data. For example, the words ‘the’, ‘cand’, and ‘have’ are frequently repeated without much sense: *I have the have the cand the have that the cand the cand the seep that the the have the seep the cand the senter the cand the seep the seep the seep the seep the cand the seep the seep the cand the*

Medium temperature: The generated text is moderately diverse, with a balance between exploration of new words and keeping learned patterns. Sentences are somewhat coherent and follow the language style, but the vocabulary is still not grammatically correct. For example,

Whim that sollied the mest and man a sold made mise this of the me wit in the man the man the sard the man the man are of you say shall my the and man the mame this om and this and le emame the sard of

High temperature: The text is more exploratory and chaotic. There is a lot of randomness, with more diversity in words and structures. However, it also increases our chances of generating nonsensical or grammatically incorrect sequences. It follows very closely the structure of Shakespearean text.

TOP-P Sampling

top_p = 0.3: With lower threshold, the sampling focuses on the most likely tokens, which gives us more predictable text generation. Thus, we should have more coherent sentences but with less variability, but I don’t notice a big difference between the two top_p values.

top_p = 0.9: Here, more tokens to be considered, and so we have an increased diversity of the generated text. This means that there is more noise introduced to the text due to randomness.

TOP-K Sampling

top_k = 40: A moderate restriction of the sampling to the top 40 tokens with the highest probabilities. We have a balance between exploring new tokens and maintaining coherence. The text should still be somewhat diverse but make more sense. For example, this sounds like something Shakespeare would say: *I have and the cait. Hood and as that make and the mone the now the now were,*

top_k = 100: With a larger k value, more tokens are considered, so there is more exploration of the vocabulary. We can have a more diverse vocabulary but more nonsensical sequences.

Considering all the samples, I observe that higher values for all parameters introduce some nonsense in a text that is already hard to understand. However, it did better on lower values, close to medium. Comparing RNN and LSTM, Ic can conclude that LSTM produced more coherent text that somehow made sense in some parts of the sequence as opposed to RNN.