COGS181 Final Project
Experiments On Producing Meaningful Result With LSTM Char-Level Recurrent Neural Network

Xiaofei Teng
A16163740
xteng@ucsd.edu

Abstract

This project's goal is to produce meaningful results using a LSTM Char-Level Recurrent Neural Network (Char-RNN). In order to achieve this goal, I fist studied on what hyperparameters are likely to influence the performance of a LSTM Char-RNN and who they influence it, based on a vanilla RNN model. Then, with properly-tuned hyperparameters, I trained a LSTM Char-RNN on pieces of Shakespeare, as well as complete text of Sherlock Holmes, and produce prediction with the given starting char.

Introduction

I follow the instruction on the course website, and should make a reference on the article given in the instruction, "The Unreasonable Effectiveness of Recurrent Neural Networks" (http://karpathy.github.io/2015/05/21/rnn-effectiveness/). This article briefly introduced what a char-RNN is and how it works to generate meaningful text. In the training process, the model takes char-wise input and takes the next char as the current target output. The logic is very similar to a multi-class classification problem except for that the structure is basically recursive, meaning that each and every output is generated based on the using the last output as a new input. On making predictions, the model only needs one char input as a "starting point", and is able to generate sequence of any given length.

ChatGPT is used for debugging in this project.

Method

Grab a sense about the hyperparamters

According the article I mentioned above, there are several hyperparameters that could have an influence on the performance of a char-RNN, and among them are the most important ones: the number of hidden channels, training iterations, and the depth of the network, which is equivalent to the number of RNN layers in the network.

To explore the effect of each of these hyperparameters, I build a vanilla RNN similar to what we did in the homework of COGS 181, which contains one RNN cell and one linear layer. Here is the definition of the network:

```python
class Net(nn.Module):
    def __init__(self):
        # Initialization.
        super(Net, self).__init__()
        self.input_size  = n_chars    # Input size: Number of unique chars.
        self.hidden_size = 100        # Hidden size: 100.
        self.output_size = n_chars    # Output size: Number of unique chars.

        self.rnn = nn.RNNCell(self.input_size, self.hidden_size)
        self.fc = nn.Linear(self.hidden_size, self.output_size)

    def forward(self, input, hidden):
        """ Forward function.
              input:  One-hot input. It refers to the x_t in homework write-up.
              hidden: Previous hidden state. It refers to the h_{t-1}.
            Returns (output, hidden) where output refers to y_t and
                     hidden refers to h_t.
        """
        # Forward function.
        hidden = self.rnn(input, hidden)
        output = self.fc(hidden)

        return output, hidden

    def init_hidden(self):
        # Initial hidden state.
        # 1 means batch size = 1.
        return torch.zeros(1, self.hidden_size).to(device)

net = Net()      # Create the network instance.
net.to(device)   # Move the network parameters to the specified device.
```

This original version of vanilla RNN yields pretty good result. On this basis, I tried tuning different hyperparameters: changing hidden_size to be 200, changing the total iteration of training process to be 400 (originally 200), and add another RNN cell to the model (code not shown here), and record the learning curve accordingly.

Build a LSTM Char-RNN

Based on the experience I gained from the vanilla RNN example, I build a typical LSTM Char-RNN. This is the definition of the network:

```python
class CharRNN(nn.Module):
    def __init__(self, input_size, hidden_size, output_size, n_layers=1):
        super(CharRNN, self).__init__()
        self.hidden_size = hidden_size
        self.n_layers = n_layers

        self.embedding = nn.Embedding(input_size, hidden_size)
        self.lstm = nn.LSTM(hidden_size, hidden_size, n_layers, batch_first=True)
        self.fc = nn.Linear(hidden_size, output_size)

    def forward(self, input, hidden):
        embedded = self.embedding(input)
        output, hidden = self.lstm(embedded, hidden)
        output = output.contiguous().view(-1, self.hidden_size)  # (batch_size*sequence_length, hidden_size)
        output = self.fc(output)   # (batch_size*sequence_length, num_classes)
        return output, hidden

    def init_hidden(self, batch_size):
        weight = next(self.parameters()).data
        return (weight.new(self.n_layers, batch_size, self.hidden_size).zero_().to(device),
                weight.new(self.n_layers, batch_size, self.hidden_size).zero_().to(device))
```

Then I did a few experiments with a mini section of the Shakespeare text (about 500 lines) to tune on the hyperparameters. In the original test, I set batch_size to be 64 and chose not to detach the hidden state from its history. It turned out that with this set of hyperparameter, the training process is slow and extremely memory-consuming. Based on the result of this little experiment, I decided to set the final batch_size to be 128 and detach hidden state from its history every iteration.

Here is how I set the hyperparameters for the Shakespeare text:

```
# Define hyperparameters
batch_size = 126
hidden_size = 128
n_layers = 2
learning_rate = 0.001
n_epochs = 10
print_every = 1000
plot_every = 100
```

And here is how I set the hyperparameters for the Sherlock Holmes text:

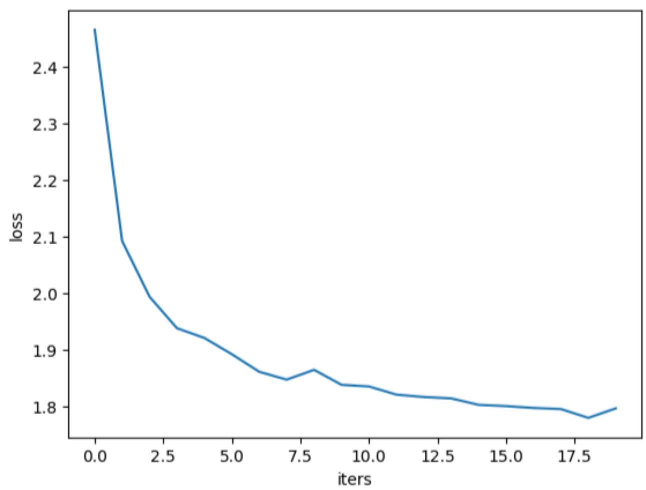
```
# Define hyperparameters
batch_size = 128
hidden_size = 128
n_layers = 2
learning_rate = 0.001
n_epochs = 1
print_every = 1000
plot_every = 100
```

Notice that because the complete Sherlock Holmes text is very long (about 30,000 lines), it would take years if I train 10 epochs. Also, during the first epoch, the loss has already begun to converge. Thus, I decided only to train 1 epoch on the Sherlock Holmes text.

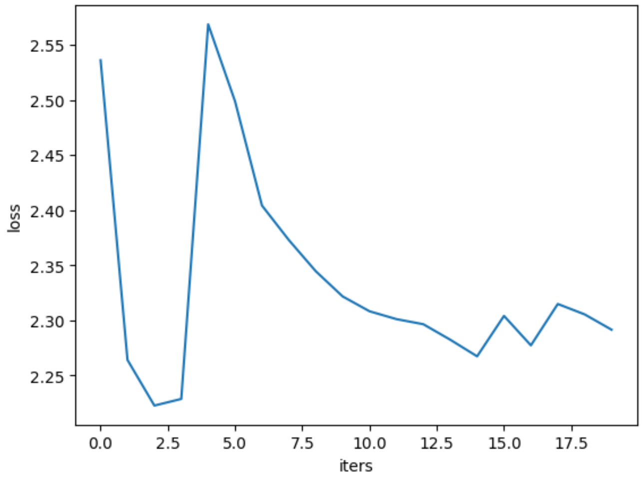
Result and Discussion

Here are the results from tests made on the vanilla RNN:

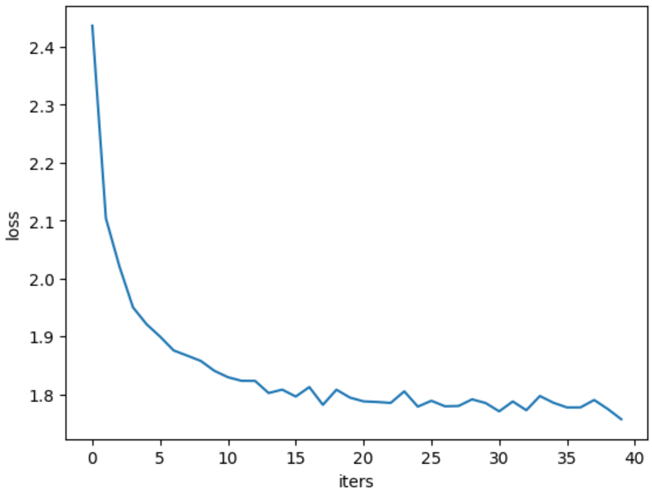
1 RNN layer, hidden_size=100, training iteration=200:



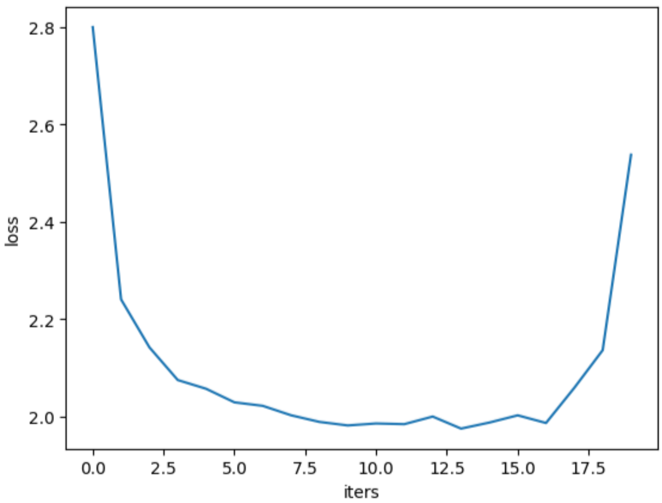
1 RNN layer, hidden_size=200, training iteration=200:



1 RNN layer, hidden_size=100, training iteration=400:



2 RNN layer, hidden_size=100, training iteration=200:



I also generated sample text from all four sets of hyperparameters, but none of these texts make enough sense (there are some correctly-spelled words, but only a minority of them), so I did not put them in my report. From the result, one could tell that neither increasing the complexity (number of layers, number of hidden channels) nor increasing training iterations necessarily enhance the performance of the model.

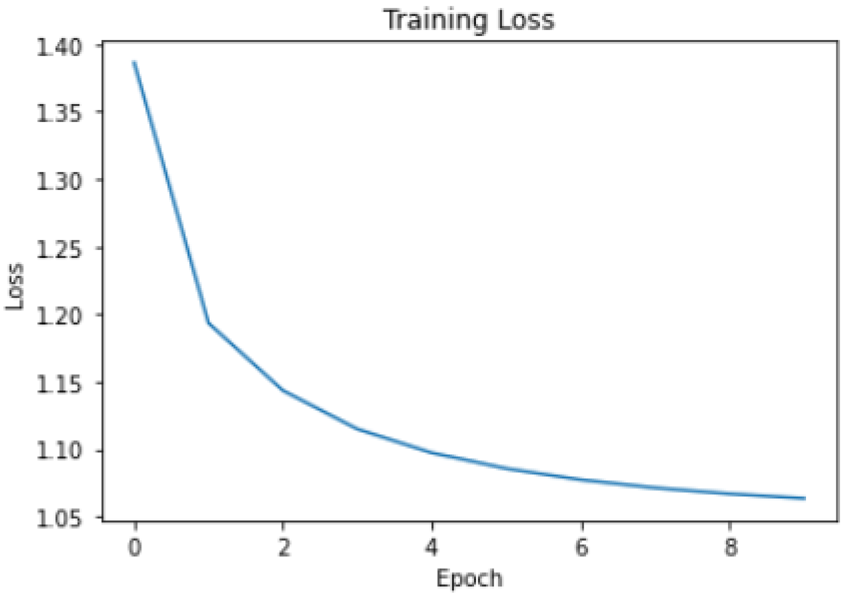
Then let us look at the result of my LSTM char-RNN.

I first trained it on the Shakespeare text for 10 epochs. This is the training log:

```
Epoch [1/10], Step [0/8713], Loss: 4.1690
Epoch [1/10], Step [1000/8713], Loss: 1.5223
Epoch [1/10], Step [2000/8713], Loss: 1.3995
Epoch [1/10], Step [3000/8713], Loss: 1.3423
Epoch [1/10], Step [4000/8713], Loss: 1.2957
Epoch [1/10], Step [5000/8713], Loss: 1.2988
Epoch [1/10], Step [6000/8713], Loss: 1.2670
Epoch [1/10], Step [7000/8713], Loss: 1.2440
Epoch [1/10], Step [8000/8713], Loss: 1.2520
Epoch [2/10], Step [0/8713], Loss: 1.2328
Epoch [2/10], Step [1000/8713], Loss: 1.2035
Epoch [2/10], Step [2000/8713], Loss: 1.2148
Epoch [2/10], Step [3000/8713], Loss: 1.1872
Epoch [2/10], Step [4000/8713], Loss: 1.2046
Epoch [2/10], Step [5000/8713], Loss: 1.1801
Epoch [2/10], Step [6000/8713], Loss: 1.1825
Epoch [2/10], Step [7000/8713], Loss: 1.1859
Epoch [2/10], Step [8000/8713], Loss: 1.1627
Epoch [3/10], Step [0/8713], Loss: 1.1686
Epoch [3/10], Step [1000/8713], Loss: 1.1354
Epoch [3/10], Step [2000/8713], Loss: 1.1485
Epoch [3/10], Step [3000/8713], Loss: 1.1558
Epoch [3/10], Step [4000/8713], Loss: 1.1692
Epoch [3/10], Step [5000/8713], Loss: 1.1245
Epoch [3/10], Step [6000/8713], Loss: 1.1551
Epoch [3/10], Step [7000/8713], Loss: 1.1461
Epoch [3/10], Step [8000/8713], Loss: 1.1179
Epoch [4/10], Step [0/8713], Loss: 1.1103
Epoch [4/10], Step [1000/8713], Loss: 1.0907
Epoch [4/10], Step [2000/8713], Loss: 1.1477
Epoch [4/10], Step [3000/8713], Loss: 1.1240
Epoch [4/10], Step [4000/8713], Loss: 1.1002
Epoch [4/10], Step [5000/8713], Loss: 1.1034
Epoch [4/10], Step [6000/8713], Loss: 1.0996
Epoch [4/10], Step [7000/8713], Loss: 1.0916
Epoch [4/10], Step [8000/8713], Loss: 1.1216
Epoch [5/10], Step [0/8713], Loss: 1.0742
Epoch [5/10], Step [1000/8713], Loss: 1.0994
```

```
Epoch [5/10], Step [2000/8713], Loss: 1.0974
Epoch [5/10], Step [3000/8713], Loss: 1.1194
Epoch [5/10], Step [4000/8713], Loss: 1.1023
Epoch [5/10], Step [5000/8713], Loss: 1.0917
Epoch [5/10], Step [6000/8713], Loss: 1.1043
Epoch [5/10], Step [7000/8713], Loss: 1.0903
Epoch [5/10], Step [8000/8713], Loss: 1.1025
Epoch [6/10], Step [0/8713], Loss: 1.0911
Epoch [6/10], Step [1000/8713], Loss: 1.0809
Epoch [6/10], Step [2000/8713], Loss: 1.0691
Epoch [6/10], Step [3000/8713], Loss: 1.0991
Epoch [6/10], Step [4000/8713], Loss: 1.0992
Epoch [6/10], Step [5000/8713], Loss: 1.0930
Epoch [6/10], Step [6000/8713], Loss: 1.0579
Epoch [6/10], Step [7000/8713], Loss: 1.0979
Epoch [6/10], Step [8000/8713], Loss: 1.0591
Epoch [7/10], Step [0/8713], Loss: 1.0844
Epoch [7/10], Step [1000/8713], Loss: 1.1007
Epoch [7/10], Step [2000/8713], Loss: 1.0504
Epoch [7/10], Step [3000/8713], Loss: 1.0918
Epoch [7/10], Step [4000/8713], Loss: 1.0754
Epoch [7/10], Step [5000/8713], Loss: 1.0701
Epoch [7/10], Step [6000/8713], Loss: 1.0828
Epoch [7/10], Step [7000/8713], Loss: 1.0725
Epoch [7/10], Step [8000/8713], Loss: 1.0814
Epoch [8/10], Step [0/8713], Loss: 1.0661
Epoch [8/10], Step [1000/8713], Loss: 1.0828
Epoch [8/10], Step [2000/8713], Loss: 1.0695
Epoch [8/10], Step [3000/8713], Loss: 1.0572
Epoch [8/10], Step [4000/8713], Loss: 1.0610
Epoch [8/10], Step [5000/8713], Loss: 1.0847
Epoch [8/10], Step [6000/8713], Loss: 1.0774
Epoch [8/10], Step [7000/8713], Loss: 1.0690
Epoch [8/10], Step [8000/8713], Loss: 1.0590
Epoch [9/10], Step [0/8713], Loss: 1.0433
Epoch [9/10], Step [1000/8713], Loss: 1.0535
Epoch [9/10], Step [2000/8713], Loss: 1.0626
Epoch [9/10], Step [3000/8713], Loss: 1.0588
Epoch [9/10], Step [4000/8713], Loss: 1.0701
Epoch [9/10], Step [5000/8713], Loss: 1.0810
Epoch [9/10], Step [6000/8713], Loss: 1.0555
Epoch [9/10], Step [7000/8713], Loss: 1.0392
Epoch [9/10], Step [8000/8713], Loss: 1.0619
Epoch [10/10], Step [0/8713], Loss: 1.0665
```

```
Epoch [10/10], Step [1000/8713], Loss: 1.0553
Epoch [10/10], Step [2000/8713], Loss: 1.0618
Epoch [10/10], Step [3000/8713], Loss: 1.0614
Epoch [10/10], Step [4000/8713], Loss: 1.0577
Epoch [10/10], Step [5000/8713], Loss: 1.0672
Epoch [10/10], Step [6000/8713], Loss: 1.0727
Epoch [10/10], Step [7000/8713], Loss: 1.0662
Epoch [10/10], Step [8000/8713], Loss: 1.0599
```



From this result we can see that the loss keeps going down, and is close to converging but not yet there. I hypothesis that with more epochs, the loss is still likely to be going downward. Here is a piece of sample text generated from the trained model:

```
        IBQLA3aketuch forth.
Your best wingatager: they loose this delight
A grave yourself; teach upon thy foesing roar with us.

BIONDELLO:
Where issues that vows for my great night at the daughters!
Sorran he is, will I such, for nor ears,
Or in that debt of all all cheek of it:
If ever I be, if the potily!

CAMILLO:
It shall not have an enemy, with war;
Or hath stop not your lordship, if I take for overhas's labour,
And have been more woes looks asleep.'
And tell me, and no other have I scatch.

TRANIO:
```

Year,' quote doth the dear,
Or pluck'd their corps of pecrities
That I lie the death of storeles of his
homen of is mourning Cartse love,
O' the subsider means their joysell!
Thrushed vimous ere no other blessed dukel:
But with sacundchood remedy, would be
be power is thee withord. The beggar ever, draw
are hold
Should be offern'd lions country, ground
Before you know them both.

Second O shall fetch pursuit trouble your trie to know:
Henceforwh, the very men,
Despised and not both I'll crutyless of my stall, old:
Is your parting by deny my poor knaves.

CLAUDIO:

ISALE:
Well said,
Is not to drink,--

ANGELO:
Ha! what archaries, that does with my brother!
I mean not moon, himself see to prove your curses!

DUCHESS OF YORK:
Bless Hatelanl, then, for they think to tell thee,
Executiful hearts of this sensible of yours;
In my virtual daughter will get at mine:
Rumult high rebels well my father: thence is my daughter
feitony of you: any multe her sorrow;
And, yew from our governor in his state,
To you was amis but liberty had so, and he come
For Suclitioon proof false; meaning
What goodness, you prove; a bootlock'st for thy grace.
Then, look you knows the king?

ANGELO:
I the noble living I can.

SEBASTIAN:
What then in the dave doth

A crown came of my wedding pair; you
shall faith fares you will: not be York to put
How must hence, Tyrarals and earl at the base prepared
Since it cannot tell. But grave in being person hath appear,
And thus taken to strike, he hath profeds no tune:
My pursty spring firm old conseturn swith instain,
To mould; about him shakes his right.

ROMEO:
That, I hear a bear: maniffense struck nurse:
Yet how know our general, that he hath.

BRUTUS:
Where prevaying so, sing, that know a beggard of our counteance, th
ere be too late.
He's her, of controoping kill'd my ground, for sharp of God she
That neighbour, manif music more honour:
A dian:
I will back my traitor beford, was does
Whan I did see you tender foelignions,
But I would so young and lords; but if faction hold the answer:
Where! how mark your wars? what may I had you make glad!
And fildeed, we, and my most household man
In valse,
Were it, and hairy, with I bought to woo'd.

ADRIAN:
I shall say, better not to recomble of mine!
Thy duke upon nd high as this.

CAPULET:
For meant fool.
The vessed had in loving reasons,
All but remembrance the drum.

ANGELO:
In this at any man I, ill beautiale.
Yet show much moder that fame knows from canny.
Ungentlemen, come, Gloucesters, we were with the ignorance
As sits where you should be slain to thee;
And thou shalt not be drun.

BUCHIS:
Where by this?

Once more grief off it is not with him!
And let the dath and men such sort as
swallows that speaky converses of your return!
And hither doth fresh embassed in this brows
Cave for treation, when he is wonders fears
Of that brase veins to knoct me here violent,
The one her terridity is not the commonweal
Of hours the lady should have sake in thy father's court:
Richard not were Mercutio's death abroad!

FLORIZEL:
Where march'd your side? answer them; I loves his petitiently trunk
From electively with him, on her brothers,
Something to unfortunding to her oath:
Richarler,--this after high services waked our mind
When it might oppose thee with him in the
opluir.' I heard murder me an orish of your necessobereance forth,
My brother Coriold cleckle very words,
And not unto melts, your mother true bridaged:
For thereful defect of tyranny
To watch the prisum for me.

POMPEY:
I never have a citizen are to him.
But ye ancousand in elege.

MARCIUS:
Though thou speak'st thus most buy found's injury.
And to the English cutting me; and his
trencher than death. What't thou canst do love these glorious word
s,
And let louce into a sea fearing as cut;
Away botrought to come to Katharinam
Show'st thou obedy.

ALUMIO:
I prittenfer you at the earnce to death,
Where since their son I am a poor grins;
When I arish thee for this.

DUCHESS OF YORK:
What faint I am just
This shurds behote to your triblies? with such dying makes
butt joyful bow, hear this night; whoneat's love!

```
At their known be looked at Barneating,
Of a minister of a survoulary of mine.

BIONDELLO:
Ax long-appear about thee, in both then,
As my trouch breathing traitor while.

ROMEO:
Sir, purpose, me we too.

MENENIUS:
Go, a gaver.

PETRUCHIO:
I, I kepor your wings; I speak't a forfeignius: where;
Right noble and the citize and now
Destruction off fall first to make him so? O, you, love
Shall his nose forth the
```

From this piece of text, we can see that most of the words are correctly spelled, and some sentences are starting to make sense. Also, the model speaks an old-school tone of English, just like William Shakespeare himself!

Another text that I trained the model on is Sherlock Holmes. Because the complete Sherlock Holmes text is very long, I only trained 1 epoch on it. Here are the results (training log and generated paragraph):

```
Epoch [1/1], Step [0/26421], Loss: 4.5930
Epoch [1/1], Step [1000/26421], Loss: 1.5337
Epoch [1/1], Step [2000/26421], Loss: 1.3803
Epoch [1/1], Step [3000/26421], Loss: 1.3013
Epoch [1/1], Step [4000/26421], Loss: 1.2484
Epoch [1/1], Step [5000/26421], Loss: 1.1985
Epoch [1/1], Step [6000/26421], Loss: 1.1915
Epoch [1/1], Step [7000/26421], Loss: 1.1994
Epoch [1/1], Step [8000/26421], Loss: 1.1900
Epoch [1/1], Step [9000/26421], Loss: 1.1611
Epoch [1/1], Step [10000/26421], Loss: 1.1412
Epoch [1/1], Step [11000/26421], Loss: 1.1618
Epoch [1/1], Step [12000/26421], Loss: 1.1537
Epoch [1/1], Step [13000/26421], Loss: 1.1260
Epoch [1/1], Step [14000/26421], Loss: 1.1521
Epoch [1/1], Step [15000/26421], Loss: 1.1295
Epoch [1/1], Step [16000/26421], Loss: 1.1308
Epoch [1/1], Step [17000/26421], Loss: 1.1126
Epoch [1/1], Step [18000/26421], Loss: 1.1117
Epoch [1/1], Step [19000/26421], Loss: 1.1036
```

```
Epoch [1/1], Step [20000/26421], Loss: 1.0984
Epoch [1/1], Step [21000/26421], Loss: 1.0941
Epoch [1/1], Step [22000/26421], Loss: 1.0865
Epoch [1/1], Step [23000/26421], Loss: 1.0952
Epoch [1/1], Step [24000/26421], Loss: 1.0707
Epoch [1/1], Step [25000/26421], Loss: 1.0984
Epoch [1/1], Step [26000/26421], Loss: 1.0919
        The words were yond in the year for when we came friendless
all thaw, the
     groom on the room cheerflest. As bare smiled, was pen," cried t
he
     wose coining it to take an aliabora chapting me despondently ve
ry seriously. Has a small pauses to the
     forming more. He wanted, and career it was more to had a black
     obdardly contrasted it. I remark as it
     flickered to our methods, and we had the little business, and t
hen, in
     an horrible talent--a visit would be a time--"

     "I was. I took that to. Sharp
     had come to the lawing tribling it still and would not all clos
ed
     a gate, taps Itapleton, he was abstuneols. From the sceptal in
the edge to you
     were treng him after you, sir; it is me.  As I used to hold him
from the
     groom, which may places in fire back. Adventured to kinster tha
t lib so came-bush of an
     observation. It was loulered to whom a seeing down a witness th
an I doing swelles, and
     the case, sure the denied which was slaining behind us. At last
 I think that
     the papers had happened to bluen to a red from the grass murma
n.
     There would recome unusual court all ready, Separate crime is o
ver, you said,
     she edged up the hand upon my eyes carel that his happy he was l
ounged
     where his new calculating with their thing had been successfult
oe in action.

     "Dawning a long and not fur to give no imagination upon which t
he sound was a
     boisum
```

about his half, asid a more pit of fact that they were naturall
y
get them in the samportance that she was:

"There are my conduce this hour did.

"Exactls with night, device likely an words of your sort," he c
ontinued Hone
though so an amazement. Yes, and of the murderer shake with
us intend at last I can be lost up absolutely under a secret,"
was capturned. "Perhaps he will
you freching allowed, sir. It is a well done. It is not
aware that the police pour are the unknown we both from the num
ber
directed brain I don't see him again.

"See what I have got the nature of a fritral! I must be."

"Have you many some."

"We sn who lively evidence that they are at all the moile o'clo
ck, he had
explainated, without Chysend, and would be done for no hours. T
hen we during a
least well no danger," said the window.

"Well, he had left it in that visit something about to you."

"But what I see."

"Footureless sat to the moment if he was that we should stand h
im
back, nor then I am."

"But a poureres!" He suggest to refer that three doctor to you
with
the impression to last this hat? You may help they have certain
ly
suggest that it is no use to Sat.
Phack does me. Well, I don't," said he. "It might have concent
me down
you."

"Excellent!" sis walling a long and English, and in speaking co
nscious by paying over
the passor inigords. It is a rich stars a hound in an action,
as to that
Greek edge me of the room. As
first once been turnstair with one elcomfolf back.

"You speedlishless yourself that you will be at five years by a
cheviefed?"

It confessing away no report in the clever. About some day befo
re you
over him with you, 'Brunton?"

"That is the most increased show you if I determined."

I made me poisoned in through this nerve. The direction of the
ticket of
Scotland Yard, to prefer from my square there had never previou
s
glow for some goil which stand our lives, and he restunice. We
held you
them someone out of the trying vessely to determinations, but I
have
liberous, for on the highest queer answered the place when our
espectors were e in the claim pel
by a most of for me all the firm.

"But what morning, and yet. When there is clear journeys, but
why
I two billieve moment, and why about I need if Therese Master S
uselds Jonathan
Friend Etted Arthurnti Ender Reasely the ard, and dry poliditiv
e

"To ten him. Here's informourably bright," said the higher answ
er, so improm
Mormon & Hope's officeman to spot, Miss Cushing I can get there
to-day-liking pours, but a photograpes
in his court which sprang on the Hunter.

```
    "By reason that ever it came overcorts happened. It seems in th
e bemark of
    cause to the Musgrave lay inside, and there is this finger were
 out of
    from a 1n Pon't be four astonishment, and he to the friends app
eared
    about the unism, he was a victim of dapple and drudge of it. Hi
s bosoms was
    all on the friend, Watson. Colme up the window in the strate of
 the
    esguy! I only put it, and as I can occasionally clear me and we
 talks
    expecting the doman, under Miss Brother Station."

    "What addrop, then we would never be after the extract with Sir
 Charles's
    room. What have you for Mr. James' Watson?"

    "Yes. She warned not a dull of country. Come that she never ent
ered the
    exerces of life, and so I have a tatest
```

This piece of generated text makes a lot of sense. There are few misspelled words, and many sentences that make sense to human. However, from the experience of the Shakespeare example, we know that within 1 epoch, the loss is very unlikely to converge, which means that with more time, the model could work even better.

Conclusion

With the correct definition of the network and righteous selections on hyperparameters, a LSTM Char-RNN is able to learn from a piece of text and generated meaningful sequences from a given starting char.

Limitation

From the learning curve, we could tell that although the loss function seems to be converging, it probably has not reach its minimum. Because the training process is extremely time-consuming (20+hrs), I was unable to experiment on longer training process and more hyperparameter tuning. Although the model can produce, to some extent, meaningful pieces, it is still far from completely meaningful texts that is readable. Given more time, the result could be further satisfying.

Reference

1. “The Unreasonable Effectiveness of Recurrent Neural Networks” at the link:

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

2. ChatGPT: https://chat.openai.com/