

HW3 Report

學號:311512015 姓名:謝元碩

一、 Standard RNN

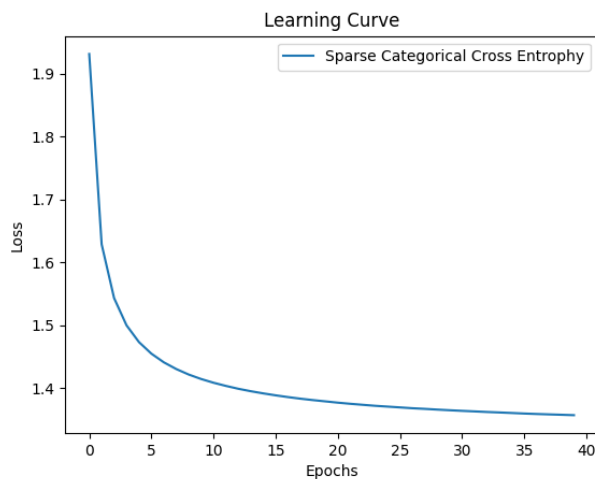
1. 訓練結果

(1) Network architecture

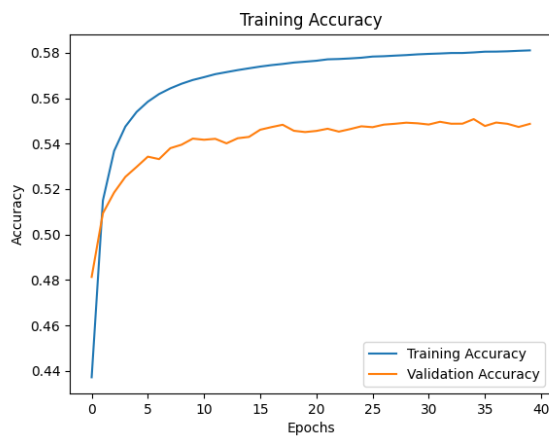
進入隱藏層前，首先對文字輸入做 word embedding 進行降維，embedding dimension 設定為 256；隱藏層部分使用一層標準簡易 RNN 並設定 neuron 單元數，經測試後使用的最佳參數為 256 個單元，以及一層 softmax 的全連接層做為輸出。Input_length 為輸入字元長度，可自行設定，根據測試後的最佳訓練結果，我們把他設定為 50。

```
Embedding(input_dim=vocab_size, output_dim=256, input_length=input_length),  
SimpleRNN(64),  
Dense(vocab_size, activation='softmax')
```

(2) Learning curve



(3) Training error rate & Validation error rate



此處使用 SparseCategoricalAccuracy 去針對訓練集和驗證集輸入去估測正確率，紀錄正確率變化曲線如上圖。結果表示 Training accuracy 為 0.5820、Validation accuracy 為 0.5520，代表 **Training error rate 為 0.418**、**Validation error rate 為 0.448**。

2. Choose 5 breakpoints during your training process to show how well your network learns through more epochs. Feed some part of your training text into RNN and show the text output.

總訓練 epoch 數為 40，因此紀錄 8、16、24、32、40 這五個 epoch 時的權重去觀察文本生成情形。

這裡使用訓練集文本中其中一段文字”We are accounted poor citizens, the patricians good.”做輸入以觀察文本生成結果，可以看見生成的文字內容皆有一定的完整性及邏輯，由於訓練時在 epoch 約為第 8 次時 loss 就已即將進入收斂，後面所訓練出來的模型成效差異性不大，才會導致此報告中顯示的輸出結果，然而在較大的 epoch 所訓練出來的模型，所生成的文字內容可讀性仍然較高，如 Give me now 和 were confess when...等。

(1) Epoch=8

```
We are accounted poor citizens, the patricians good.

EDGAR:
Seeked throat horse--
You should died thee order I betray rudines.

COUNTESS:
Will sid fight ready; indeed in him's't, all 'ten Roman'd
My mocknows
Whis is before she got you abide a musid's
state for your mind, go have assembled, I must rull'd in France in this time!

MOCTOLES:
Youk her amb
```

(2) Epoch=16

```
We are accounted poor citizens, the patricians good. This is her: who ever: 'tis a foot.
This ignorance daughterers contenden and be pupper?

EDGAR:
Within the Lord, thou wilt very poor on his decion immood, hath beeketh.

Post:
Lord Somewithant i shaled;
And do we will forth;
Let stands,
That not keep their forbank his garsed
In laten bears the art,
```

(3) Epoch=24

```
We are accounted poor citizens, the patricians good.

BARDOLPH:
Tut been my love
Living the land, thou'lt some sweet banquet you worn and trear;
First, will you will needs seem,
Prisone oft;
You call beheld, no other, o' such a becausion:
Against thy heart we full ug?

Nurse:
O my lord to ordes speak not amasure out thousand tainter gave and desert
```

(4) Epoch=32

```
We are accounted poor citizens, the patricians good.

VALENTINE:
Are a mountain
Shall
should not come
That
I bed thy loathful and hand of my anger nege miseries hath the gait,
Gatest here.
hapit thy purse, your commot heart,
She hath say no more
As she made markinead.

HORATIO PETERONE:
His innoce I to his person.

BELARIUS:
Good Master Barndellow,
```

(5) Epoch=40

```
We are accounted poor citizens, the patricians good.

Sicheman:
Of my Lord our thick is no chief-spom healthy, ring,
Yet the duke.

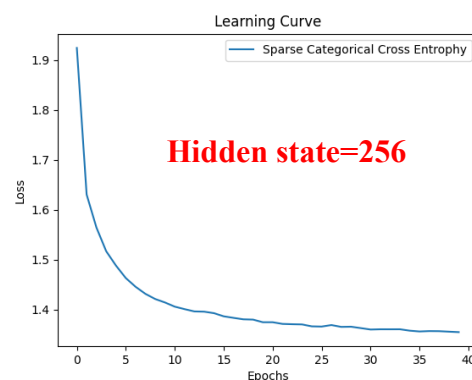
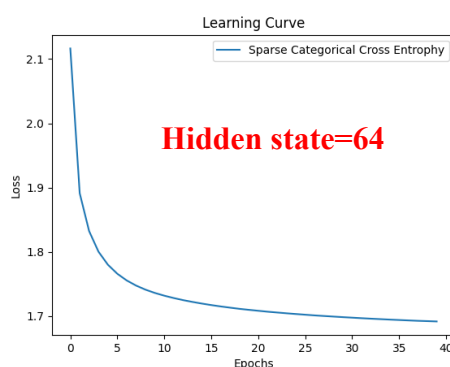
GRATIANO:
He for the sorted or it shame.
And you, Perivish, double-matters in the pay.
Look tops to be Freeth. If I think we well as masterdest may be a
hard it is
wouldst thou brow.
Give me now,
Were confess when the pa
```

3. Compare the results of choosing different size of hidden states and sequence length by plotting the training loss vs. different parameters.

(1) 不同大小的 hidden state

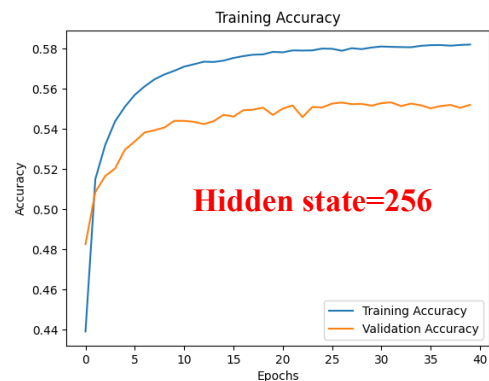
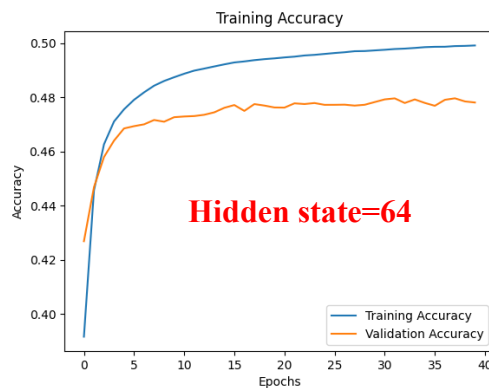
在超參數相同情形進行比較，sequence length 皆設為 50，hidden state 以 64 和 256 兩種進行訓練並比較結果。

Loss 變化如下，最後 hidden state=64 的 loss 為 1.6914，高於 hidden state=256 的 loss:1.3546，可以猜測較少的隱藏層單元數會直接影響 loss 的下降程度，甚至影響訓練結果。



正確率變化曲線如下，hidden state=64 的訓練集正確率=0.4991、驗證集正確率

=0.4781，hidden state=256 的訓練集正確率=0.5820、驗證集正確率=0.5520，可見當使用較少 neuron 數所訓練的模型，正確率也因此降低。



最後，根據前述訓練結果而去觀察文本生成結果如下圖。

Hidden state=64

```

JULIET:
The condsoUICHELLOWINO:
But with may sir, thou purno:
A letener age
Swere so lehe now!

POTUMPEROGPO:
Ay lehlo Frath werriporo.

SARANELLO:
Go waw.

EROSE: your mays
I deidy,
But that:
To oum, thou loot,
I wis all somen Live fear, your laine, singt us you!
It henewam unnove:
Late is roun, ang mork
    
```

Hidden state=256

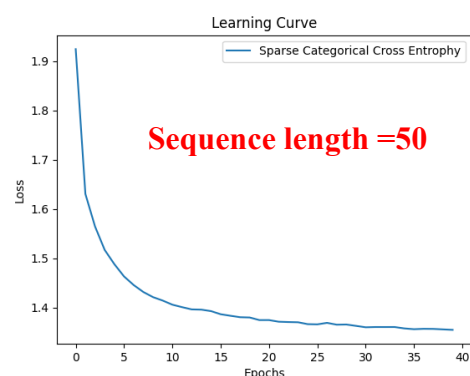
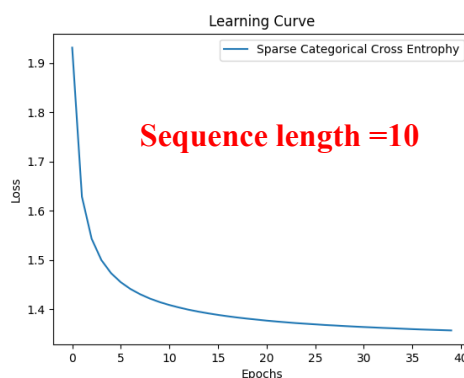
```

JULIET:
Our jery with Lenence:
These:
Ay, good Humblasy: none Amis:
Mennca, conCERS:
Are youthSS:
She wtlliclacullownessale I:
That Stain Card:
Ay, as I:
I may Lans Pyord:
Fee, Godimageab:
He is and Gornen:
Have George:
For few:
Indeence:
Franor:
No tremain:
O, none:
Woe!'
Andight:
Drawhon:
Prospursiba fr
    
```

(2) 不同大小的 sequence length

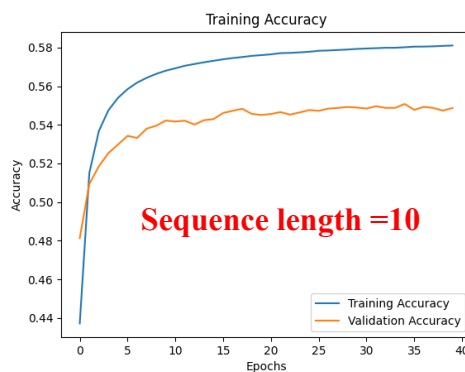
在超參數相同情形進行比較，hidden state 皆設為 256，sequence length 以 10 和 50 兩種進行訓練並比較結果。

Loss 變化如下，最後 sequence length=10 的 loss 為 1.3572，高於 sequence length=50 的 loss:1.3546，兩者在 loss 的差異性較小。



正確率變化曲線如下，sequence length=10 的訓練集正確率=0.5811、驗證集正確

率=0.5487，sequence length=50 的訓練集正確率=0.5820、驗證集正確率=0.5520，兩者差異同樣不大，故猜測 sequence length 對於數據上的訓練結果較沒有影響。



文本生成結果如下，可以觀察到 sequence length=10 所生成的文字較有段落感，不同角色的對話內容完整；sequence length=50 的文字雖單字大多正確，但無段落感，且每個句子結尾大多為冒號，難以構成文章，因此推定合適長度的 sequence length 才能夠真正訓練出好的文本生成模型。

Sequence length = 10

```

JULIET:
Ay, one-seemitary, this bold, who hear now? O.
Dost Wallind thy protest so for allow looks?

LUCIUS:
Well,
And wast
Hiscouzed to die or the fast in.

PERIT:
I roub me, have they marken it,
To that contends:
Where comes!
What you no.

GOOBLONDOGARO:
It woulderes, go come?

SOMENG:
No, on'r for conf
  
```

Sequence length = 50

```

JULIET:
Our jery with Lenence:
These:
Ay, good Humblasy: none Amis:
Mennca, conCERs:
Are youthSS:
She witliclacullownnessale I:
That Stain Card:
Ay, as I:
I may Lans Pyord:
Fee, Godimageab:
He is and Gormen:
Have George:
For few:
Indeence:
Franor:
No tremain:
O, none:
Woe!'
Andight:
Drawhon:
Prospurstba fr
  
```

二、 LSTM

1. 訓練結果

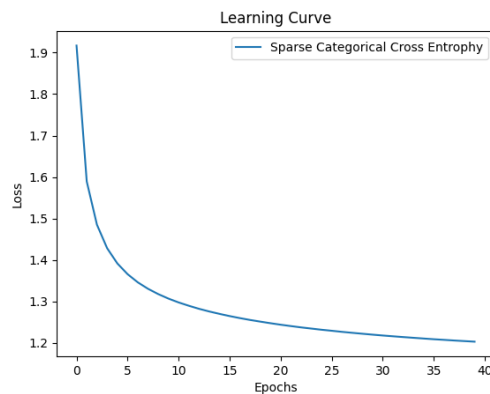
(4) Network architecture

進入隱藏層前，首先對文字輸入做 word embedding 進行降維，embedding dimension 設定為 256；隱藏層部分使用一層 LSTM 並設定 neuron 單元數，經測試後使用的最佳參數為 256 個單元，以及一層 softmax 的全連接層做為輸出。Input_length 為輸入字元長度，可自行設定，根據測試後的最佳訓練結果，我們把他設定為 50。

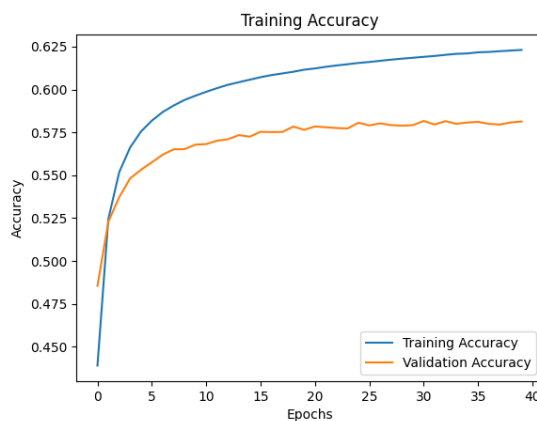
```

Embedding(input_dim=vocab_size, output_dim=256, input_length=input_length),
LSTM(256),
Dense(vocab_size, activation='softmax')
  
```

(5) Learning curve



(6) Training error rate & Validation error rate



此處使用 `SparseCategoricalAccuracy` 去針對訓練集和驗證集輸入去估測正確率，紀錄正確率變化曲線如上圖。結果表示 Training accuracy 為 0.6231、Validation accuracy 為 0.5813，代表 **Training error rate 為 0.3769**、**Validation error rate 為 0.4187**。

2. Choose 5 breakpoints during your training process to show how well your network learns through more epochs. Feed some part of your training text into RNN and show the text output.

總訓練 epoch 數為 40，因此紀錄 8、16、24、32、40 這五個 epoch 時的權重去觀察文本生成情形。

這裡使用訓練集文本中其中一段文字 "We are accounted poor citizens, the patricians good." 做輸入以觀察文本生成結果，生成的文字內容同樣呈現出一定的完整性和邏輯，這可能是因為在訓練的早期階段（大約在第 8 個 epoch），模型的損失函數已經趨於收斂。後續的訓練並沒有顯著提升模型的性能，因此導致了報告中展示的輸出結果。然而，在更大的 epoch 數下訓練的模型生成的文字內容仍然具有較高的可讀性，例如 Give me now 和 were confess when... 等例子。

(1) Epoch=8

(5) Epoch=40

```
We are accounted poor citizens, the patricians good.

PROSPERO:
Well, I'll think upon them,
which titles trot: I fought you are now, redeem'd, the
eye feigning, Musician:
Why start his rout.

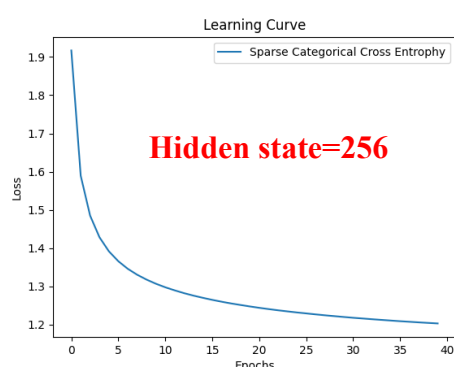
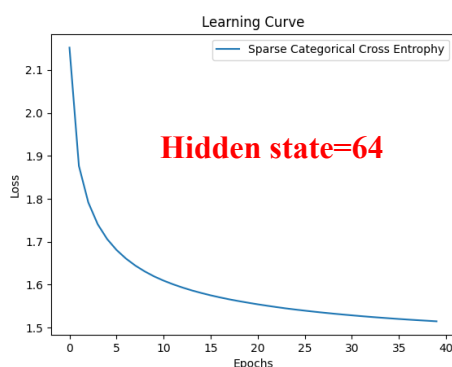
PRINCE EDWARD:
Against thy bed
cockmencate:
For Ghoundest glace, levied Milany! why, to play in tribute south.
My tongue of these be for reason stoop take your
```

3. Compare the results of choosing different size of hidden states and sequence length by plotting the training loss vs. different parameters.

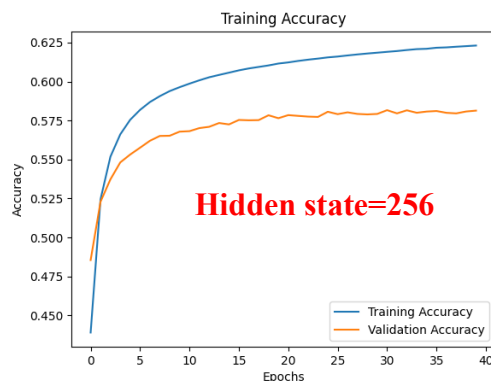
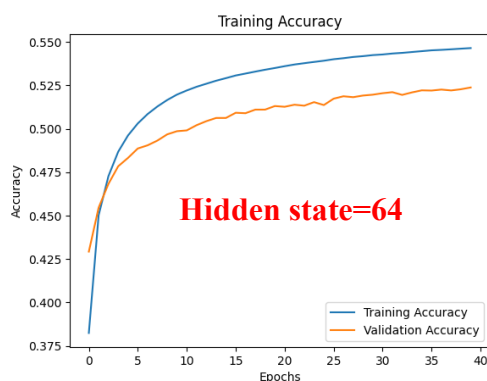
(3) 不同大小的 hidden state

在超參數相同情形進行比較，sequence length 皆設為 50，hidden state 以 64 和 256 兩種進行訓練並比較結果。

Loss 變化如下，最後 hidden state=64 的 loss 為 1.5148，高於 hidden state=256 的 loss:1.2031，可以猜測較少的隱藏層單元數會直接影響 loss 的下降程度，甚至影響訓練結果。



正確率變化曲線如下，hidden state=64 的訓練集正確率=0.5465、驗證集正確率=0.5238，hidden state=256 的訓練集正確率=0.6231、驗證集正確率=0.5813，可見當使用較少 neuron 數所訓練的模型，正確率也因此降低。



最後，根據前述訓練結果而去觀察文本生成結果如下圖。可以明顯觀察到 hidden state=64 的文字錯誤率相當大，即使訓練結果擁有一定的 loss 與正確率，hidden state=256 的文字結果同樣看不出單字，只有隱約看見句子與句子間的段落性。

```
JULIETLADOLBYBNABUGLEBNANCKYM:
VLUMYVo KNovK?
ItH?-DAMBISSLAUDICUS
XIFFFLIDVESS:
VocmYM?-GLAMUBLLUSILLLUSSLANTWSKY:WAVANVWAMLWKALISCULLANNdYGHASHMANTIWSK:
Du autTwZhZMAMGLUMISSL:
VaYNQE:-SlivK?-tujro?'a havKQUJYNNA
PHMYLLAWAMMANTW3LIWAVYML?FFNVIVLUSCAMLUSBYNBULCONBYSWAMWNKALLANTIDLUELM:
QUWVICUGBANBU
```

Hidden state=64

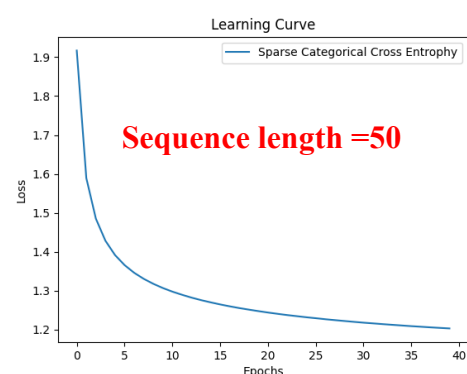
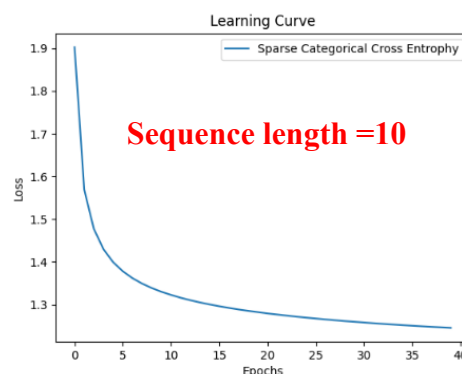
```
JULIETXAZXXrZQULIn-QZpatcky:
klivale,
ThuckUrte-prichly oftre:
Gecuck:
kepute
Alok:.
Allack:--'thlHeStentrerky, futhe LUTHLLLXAXyVX:
ke kite:
klady,
NateNer.
Nurrck, kputy
Ofk EvXdeTenvrain?
NusI,? kpalle?
WeSery: flicefutenk tufe-gHesIn,--
Lacpterveltertreekive
Therivtry, CyquHQULEZAXXXTERCASXkdI:
klack
```

Hidden state=256

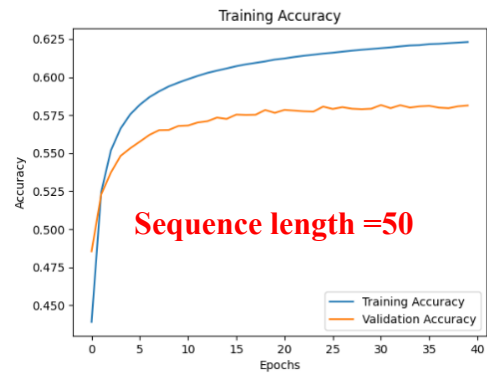
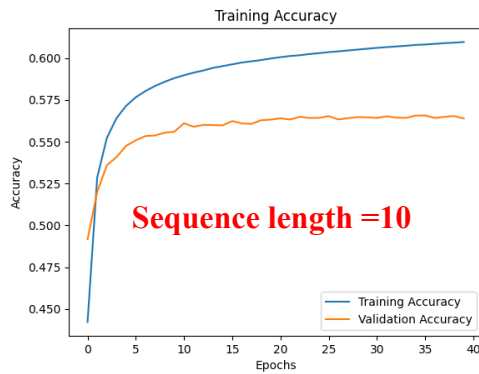
(4) 不同大小的 sequence length

在超參數相同情形進行比較，hidden state 皆設為 256，sequence length 以 10 和 50 兩種進行訓練並比較結果。

Loss 變化如下，最後 sequence length=10 的 loss 為 1.2457，高於 sequence length=50 的 loss:1.2031，兩者在 loss 的差異性較小，然而如同 Standard RNN 結果，sequence length 較小時的 loss 稍微小一些。



正確率變化曲線如下，sequence length=10 的訓練集正確率=0.6096、驗證集正確率=0.5639，sequence length=50 的訓練集正確率=0.6231、驗證集正確率=0.5813，兩者差異同樣不大，故猜測 sequence length 對於數據上的訓練結果影響較小。



文本生成結果如下，可以觀察到 sequence length=10 所生成的文字較有段落感，不同角色的對話內容完整；sequence length=50 的單字大多錯誤且無段落感，難以構成文章，因此同樣推定合適長度的 sequence length 才能夠真正訓練出好的文本生成模型，不宜太短也不宜太長。

```
JULIET:
You privorest:
No, All you sweer: I part Nort:
Befor:
And officer's Crave:
But willester:
Natur:
Ay, is anst hire Sovesters:
For, Ophey:
I pralssiner off-loo far so satiga:
I do Lord HOTS:
Well,
Morer:
What Ofta:
The still:
Why brood Shepherd:
What Mart:
'Tis, fear Etwad:
More Pandanorswer:
A makal
```

Sequence length = 10

```
JULIETLADOLBYBNABUGLELBNNANCKYM:
VLUMYVo KNovK?
Ith?-DAMBISSLAUDICUS
XIFFFLIDVESS:
VocmYM?-GLAMUBLLUSILLUSSLANtWSKY:WAVANVWAMLWKALISCULLANNdYGHASMANtIWSK:
Du autTwZhZMAMGLUMISSL:
VaYNQE:-SlivK?-tujro?'a havKQUJYNNA
PHMYLLAWAMMANtW3LIWAVYML?FFNVIVLUSCAMBLUSBYNBULCONBYSWAMWWKMALLANTIDLUVELM:
QUWVICUGBANBU
```

Sequence length = 50

三、 比較 Standard RNN 和 LSTM 結果差異

1. 標準 RNN 和 LSTM 訓練結果比較：

(1) 損失函數及正確率：

LSTM 在訓練和驗證準確率上優於標準 RNN，顯示出 LSTM 在捕捉長期依賴關係方面的優勢。標準 RNN 在處理長序列時容易面臨梯度消失或梯度爆炸的問題，而 LSTM 通過其門控機制有效地解決了這一問題。

(2) 文本生成效果：

LSTM 生成的文字內容相對更有邏輯性和完整性，這可能是由於 LSTM 能夠更

好地捕捉上下文信息，有助於生成更連貫的語句。相對地，標準 RNN 在長序列的生成上可能出現信息遺失的情況，導致生成的文本內容缺乏一致性。

(3) 訓練過程觀察：

LSTM 的學習曲線相對平滑，顯示出訓練過程更加穩定。相較之下，標準 RNN 的學習曲線可能較不穩定，表現在訓練和驗證準確率的震盪。

2. 超參數比較：

(1) 隱藏層大小 (Neuron 單元數)：

在標準 RNN 和 LSTM 中，增加隱藏層的單元數都對模型性能有正面的影響。然而，LSTM 在隱藏層大小增加時的性能提升較為明顯，這表明 LSTM 能夠更充分地利用更多的參數信息。

(2) 序列長度：

標準 RNN 和 LSTM 對序列長度的變化相對響應較小。然而，觀察文本生成結果，我們發現在適當的序列長度下，模型生成的文本內容更有邏輯性，而過長或過短的序列可能導致模型難以捕捉上下文關係。

總結來看，LSTM 在處理長期依賴性上的優越性使其在文本生成任務中表現更為出色。透過 gate 機制，LSTM 能夠更好地捕捉長距離的相依性，然而在文本生成中，標準 RNN 所生成的文本更具有上下文一致性，而 LSTM 的結果相對容易因為超參數的設計導致可讀性下降以及單字生成錯誤等情形。

四、 Use RNN or LSTM to generate some words by priming the model with a word related to your dataset. Priming the model means giving it some input text to create context and then take the output of the RNN. For example, use "JULIET" as the prime text of Shakespeare dataset and run the model to generate 10 to 15 lines of output.

1. 使用 SimpleRNN 生成之文字結果

```
JULIET:
Ay, one-seemitary, this bold, who hear now? O.
Dost Wallind thy protest so for allow looks?

LUCIUS:
Well,
And wast
Hiscouzed to die or the fast in.

PERIT:
I roub me, have they marken it,
To that contends:
Where comes!
What you no.

GOOBLONDOGARO:
It woulderes, go come?

SOMENG:
No, on'r for conf
```

2. 使用 LSTM 生成之文字結果

```
JULIET:
You privorest:
No, All you sweer: I part Nort:
Befor:
And officer's Crave:
But willester:
Natur:
Ay, is anst hire Sovesters:
For, Ophey:
I praisiner off-loo far so satiga:
I do Lord HOTS:
Well,
Morer:
What Ofta:
The still:
Why brood Shepher:
What Mart:
'Tis, fear Etwad:
More Pandanorswer:
A makal
```

3. 兩種網路架構之生成結果討論

(1) 句子連貫性和語法：

LSTM 生成的文本具有較佳的連貫性和語法，句子更有結構，並且遵循邏輯流，更容易理解。

(2) 上下文相關性：

SimpleRNN 生成的文本較缺乏明確的上下文相關性，短語之間轉換突兀。另一方面，LSTM 生成的文本在每個部分內保持了更一致的上下文。

(3) 訓練效果：

相對於 SimpleRNN，LSTM 在捕捉序列數據中的長期依賴性方面效果更好。LSTM 生成的文本可能受益於這種能力，產生更具上下文相關性和連貫性的輸出。

(4) 句子長度：

SimpleRNN 生成的文本包含較短的句子和短語，這可能導致缺乏上下文和連貫性。LSTM 生成的文本有更長且結構更好的句子。

(5) 標點符號的使用：

LSTM 生成的文本更恰當地使用標點符號，有助於提高可讀性。SimpleRNN 生成的文本標點使用不一致，影響整體清晰度。

整體來說，與 SimpleRNN 相比，LSTM 生成的文本表現出更好的語言連貫性、語法和上下文相關性。這與 LSTM 架構在處理序列數據和捕捉長期依賴性的已知優勢有著一定的關聯性。