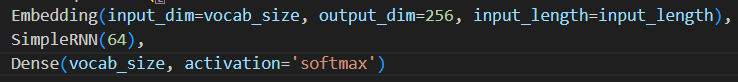
HW3 Report

學號:311512015 姓名:謝元碩

1. **Standard RNN**
   * + 1. 訓練結果
2. 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。



1. Learning curve

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述

1. Training error rate & Validation error rate 一張含有 文字, 圖表, 繪圖, 行 的圖片

   自動產生的描述

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**。

* + - 1. 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

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. Epoch=16

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. Epoch=24

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. Epoch=32

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

1. Epoch=40

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* + - 1. 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=64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。

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述

**Hidden state=256**

**Hidden state=64**

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。

一張含有 文字, 圖表, 行, 繪圖 的圖片

自動產生的描述 一張含有 文字, 圖表, 繪圖, 行 的圖片

自動產生的描述

**Hidden state=256**

**Hidden state=64**

最後，根據前述訓練結果而去觀察文本生成結果如下圖。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述 一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

**Hidden state=256**

**Hidden state=64**

1. 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。

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述

**Sequence length =50**

**Sequence length =10**

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。

一張含有 文字, 圖表, 繪圖, 行 的圖片

自動產生的描述 一張含有 文字, 圖表, 繪圖, 行 的圖片

自動產生的描述

**Sequence length =50**

**Sequence length =10**

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。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述 一張含有 文字, 螢幕擷取畫面, 字型 的圖片

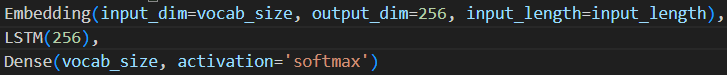
自動產生的描述

**Sequence length =50**

**Sequence length =10**

1. **LSTM**
   * + 1. 訓練結果
2. 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。



1. Learning curve

一張含有 文字, 圖表, 螢幕擷取畫面, 行 的圖片

自動產生的描述

1. Training error rate & Validation error rate一張含有 文字, 繪圖, 圖表, 螢幕擷取畫面 的圖片

   自動產生的描述

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**。

* + - 1. 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

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. Epoch=16

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. Epoch=24

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. Epoch=32

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* 1. Epoch=40

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* + - 1. 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1.5148，高於hidden state=64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。

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述 一張含有 文字, 圖表, 螢幕擷取畫面, 行 的圖片

自動產生的描述

**Hidden state=256**

**Hidden state=64**

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。

一張含有 文字, 圖表, 繪圖, 行 的圖片

自動產生的描述 一張含有 文字, 繪圖, 圖表, 螢幕擷取畫面 的圖片

自動產生的描述

**Hidden state=256**

**Hidden state=64**

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。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

**Hidden state=64**

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

**Hidden state=256**

1. 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。

一張含有 文字, 螢幕擷取畫面, 圖表, 行 的圖片

自動產生的描述 一張含有 文字, 圖表, 螢幕擷取畫面, 行 的圖片

自動產生的描述

**Sequence length =50**

**Sequence length =10**

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。

一張含有 文字, 圖表, 行, 繪圖 的圖片

自動產生的描述 一張含有 文字, 繪圖, 圖表, 螢幕擷取畫面 的圖片

自動產生的描述

**Sequence length =50**

**Sequence length =10**

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。

一張含有 文字, 螢幕擷取畫面, 字型, 功能表 的圖片

自動產生的描述 一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

**Sequence length =10**

**Sequence length =50**

1. **比較Standard RNN和LSTM結果差異**
   * + 1. 標準 RNN 和 LSTM 訓練結果比較：
2. 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。

1. 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。

1. 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。

* + - 1. 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。

1. 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。

1. **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

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* + - 1. 使用LSTM生成之文字結果

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

* + - 1. 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。

1. 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。

1. 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。

1. 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。

1. 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。