

如果你看到不一樣的結果，不用擔心，你沒有做錯任何事，純粹是神經元的隨機初始化影響了最終的分數。

改善資料

有一種小技巧可以擴展資料組的大小而不需要添加任何新歌曲，稱為資料窗口化 (*windowing*)。現在，我們將每一首歌裡面的每一行當成單獨的一行來讀取，然後轉換成輸入序列，就像圖 8-2 所展示的那樣。雖然人類是逐行閱讀歌曲來聆聽押韻和韻律，但是模型不需要這樣，尤其是在使用雙向的 LSTM 時。

因此，我們可以將每一行都視為一首歌、一個連續的文本，而不是取得「In the town of Athy, one Jeremy Lanigan」這一行，處理它，再移到下一行（「Battered away till he hadn't a pound」）再處理它。我們可以在文本裡建立一個有 n 個單字的「窗口」，然後將窗口往後移動一個字，來取得下一個輸入序列（圖 8-9）。

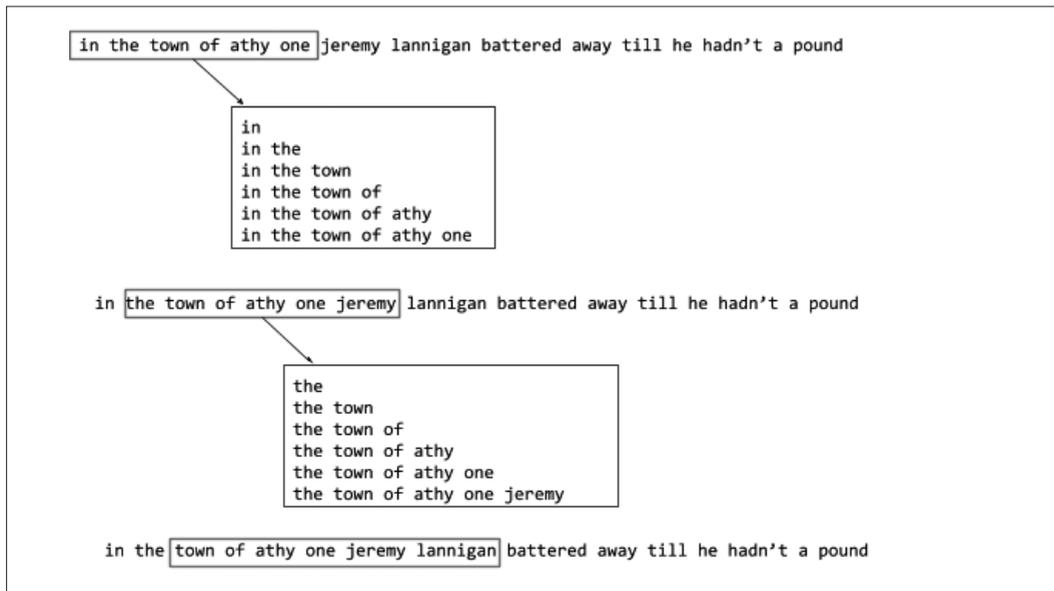


圖 8-9 移動的單字窗口

在這裡，我們可以產生多很多的訓練資料，將窗口滑過整個文本語料庫會產生 $(number_of_words - window_size) \times window_size$ 個可以用來訓練的輸入序列。

程式很簡單，在載入資料時，我們用語料庫裡面的單字動態建立它們，而不需要將每一行歌曲拆成「句子」：

```
window_size=10
sentences=[]
alltext=[]
data = open('/tmp/irish-lyrics-eof.txt').read()
corpus = data.lower()
words = corpus.split(" ")
range_size = len(words)-max_sequence_len
for i in range(0, range_size):
    thissentence=""
    for word in range(0, window_size-1):
        word = words[i+word]
        thissentence = thissentence + word
        thissentence = thissentence + " "
    sentences.append(thissentence)
```

在這個例子裡，因為我們不再使用句子，而是建立大小和窗口一樣的序列，所以 `max_sequence_len` 是窗口的大小，我們讀取整個檔案，轉換成小寫，再將它拆成單字陣列，然後遍歷單字，並且用「目前的索引」到「目前的索引加上窗口大小」之間的每一個單字來製作句子，將每一個新創的句子加入句子陣列。

額外的資料讓每一個 `epoch` 的訓練都緩慢許多，但是結果有很大的改善，而且生成的文本淪落成胡言亂語的速度慢得多。

這是讓我驚艷的範例——尤其是最後一行！

```
you know nothing, jon snow is gone
and the young and the rose and wide
to where my love i will play
the heart of the kerry
the wall i watched a neat little town
```

你可以試著調整許多超參數，改變窗口大小可以改變訓練資料的數量——較小的窗口會產生較多資料，但是一個標籤被分配到的單字比較少，所以如果你把它設得太小，你就會得到荒謬的詩歌。你也可以改變 `embedding` 的維數、`LSTM` 的數量，或用來訓練的詞彙表的大小。由於百分比準確度不是最好的指標（因為我們會比較主觀地判斷詩歌的「合理」程度），所以沒有硬性的規則可以用來判斷模型是「好」還是「不好」。

例如，當我嘗試窗口大小 6，並將 embedding 的維數加到 16，將 LSTM 的數量從窗口大小（6）變成 32，並且提高 Adam 優化函數的學習率時，我得到很棒且平滑的學習曲線（圖 8-10），而且有些詩開始變得更有意義了。

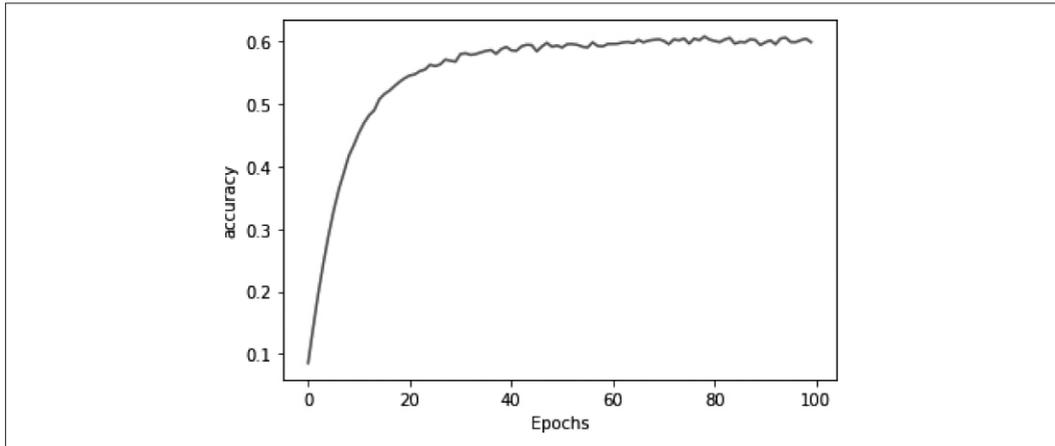


圖 8-10 使用調整過的超參數得到的學習曲線

將「sweet jeremy saw dublin」當成種子（記住，在種子裡面的所有單字都在語料庫裡）來使用時，我得到這首詩：

*sweet jeremy saw dublin
whack fol
all the watch came
and if ever you love get up from the stool
longs to go as i was passing my aged father
if you can visit new ross
gallant words i shall make
such powr of her goods
and her gear
and her calico blouse
she began the one night
rain from the morning so early
oer railroad ties and crossings
i made my weary way
through swamps and elevations
my tired feet
was the good heavens*