

陳婕儀

107403504

資管三A

## 第二次作業-RNN

連結：[https://colab.research.google.com/drive/1zB2U4ArM1zr\\_YI2COWJwo1vqfV-rDfHq?usp=sharing](https://colab.research.google.com/drive/1zB2U4ArM1zr_YI2COWJwo1vqfV-rDfHq?usp=sharing)

Text generation 感覺可以用來完成電子商務的作業，爬一堆相關的文就可以完成一篇作業，認真考慮剩下的作業要不要這樣完成，雖然產生出來的字句之間還是有點生硬，發現找的文章內無相關的字出現時他會出錯，所以如果真的要靠他完成作業，可能給他很多的資料才行。

股票預測一開始以為會很難，後來還是很努力的試試，發現其實網上有不少的相關程式碼，大家都還蠻希望可以預測出股票的走向諗，一堆不同的訓練方式或是資料的處理方式，每種方式感覺都還不錯，不過有些比較好的模型有點複雜，但我只會寫最簡單的模型，所以就沒去參考一些很複雜模型的寫法，這次作業也很有趣，跟過去的程式作業相比，這門課的作業讓我比較有動力完成它。

## • Text generation

Colab Notebooks - Google 雲端硬碟

「TextGeneration-107403504.ipynb」的副本 - Colaboratory

大學部第二次作業-RNN | 國立中央大學「新 ee-class 系統」

「TextGeneration-107403504.ipynb」的副本 ☆

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

+ 程式碼 + 文字

RAM 磁碟

編輯

```
# 作業之一就是試試看其他本小說

book = ""
with open("/content/Qiong Yao.txt", "r", encoding="utf8") as file:
    for line in file:
        book += line

book_length = len(book)
unique_words = set(book)
print(f"瓊瑤煙雨濛濛共有 {book_length} 字詞")
print(f"包含了 {len(unique_words)} 個獨一無二的字 (含標點符號)\n")
print(book[0:500])
```

瓊瑤煙雨濛濛共有 157411 字詞  
包含了 2815 個獨一無二的字 (含標點符號)

《二〇一六年十一月四日版》  
《好讀書櫃》經典版  
第一章  
又到了這可厭的日子，吃過了晚飯，我悶悶的坐在窗前的椅子裡，望著窗外那綿綿密密的細雨。屋檐下垂著的電線上，掛著一串水珠，晶瑩而透明，像一條珍珠項鍊。在那圍牆旁邊的芭蕉樹上，水滴正從那闊大的葉。  
「依萍，你還沒有去嗎？」  
媽從廚房裡跑了出來，她剛剛洗過碗，手上的水還沒有擦乾，那條藍色滾白邊的圍裙也還繫在她的腰上。  
「我就要去了。」我無可奈何的說，在屋角裡找尋我的雨傘。  
「到了『那邊』，不要和他們起衝突才好，告訴你爸爸，房租不能再拖了，我們已經欠了兩個月——」  
「我知道，不管用什麼方法，我把錢要來就是了！」我說，仍然在找尋我的傘。  
「你的傘在壁櫥裡。」媽說，從壁櫥裡拿出了我的傘，交給了我，又望了望天，低聲的說：「早一點回來，如果拿到了錢，就坐三輪車回來吧！雨要下大了。」  
我拿著傘

### 3. 資料前處理

文字前處理有一堆方法、作法:

- 切字

執行中 (已持續 1 分鐘 5 秒) Cell > fit() > \_call\_\_() > \_call() > \_call\_\_() > \_call\_flat() > call() > quick\_execute()

colab.research.google.com

HW2 - Google 雲端硬碟

StockRNN-107403504.ipynb - Colaboratory

「TextGeneration-107403504.ipynb」的副本 - Colabor...

M 大學部第二次作業-RNN | 國立中央大學「新 ee-class...

「TextGeneration-107403504.ipynb」的副本

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

留言 共用 編輯

+ 程式碼 + 文字

```

print(f"Model input shape : {input_example.shape}")
print(f"Model output shape : {predict_example.shape}")
print(f"Model target shape : {target_exaple.shape}")

Model input shape : (64, 20)
Model output shape : (64, 20, 1331)
Model target shape : (64, 20)

[33] print("原本的中文字序列:")
[print(index_2_word[ind],end="") for ind in input_example[0].numpy()]
print()
print("-"*40)
print("輸入尚未訓練的model後獲得:")
print()

predict_words = tf.math.argmax(predict_example[0],-1)
[print(index_2_word[ind],end="") for ind in predict_words.numpy()]
print()

原本的中文字序列:
了,謝謝老天!」她興奮的去端那杯牛,又要

輸入尚未訓練的model後獲得:

秘璧盤盤眉眉拚挨眉由由修修顯顯顯拚取

```

5. 制定訓練計畫並訓練

- [sparse\\_categorical\\_crossentropy](#) V.S. [categorical\\_crossentropy](#)

```

# categorical_crossentropy
y_true = [[0, 1, 0], [0, 0, 1]]
y_pred = [[0.05, 0.95, 0], [0.1, 0.8, 0.1]]
assert loss.shape == (2,)

```

執行中 (已持續 51 秒) Cell > fit() > \_\_call\_\_() > \_call() > \_\_call\_\_() > \_call\_flat() > call() > quick\_execute()

...

colab.research.google.com

HW2 - Google 雲端硬碟

「TextGeneration-107403504.ipynb」的副本 - Colaboratory

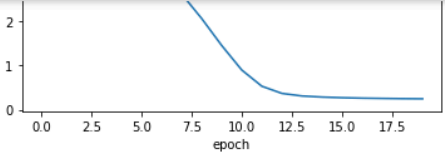
「TextGeneration-107403504.ipynb」的副本

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

留言 共用 編輯

RAM 磁碟

+ 程式碼 + 文字



```

[37] after_train_predictions = model(input_example)
after_sampled_indices = tf.argmax(after_train_predictions[0],1)

print("原本的中文序列:")
[print(index_2_word[ind],end="") for ind in input_example[0].numpy()]
print()
print("-"*40)
print("輸入進訓練後的model後獲得:")
print()

[print(index_2_word[ind],end="") for ind in after_sampled_indices.numpy()]
print()

```

原本的中文序列：  
了，謝謝老天！」她興奮的去端那杯牛，又要

輸入進訓練後的model後獲得：

，他謝老天！」她興奮的去端那杯牛，又要笑

7. 做預測

temperature = 0.01

temperature = 0.2

temperature = 0.4

執行中 (已持續 2 分鐘 34 秒) Cell

colab.research.google.com

HW2 - Google 雲端硬碟

「TextGeneration-107403504.ipynb」的副本 - Colaboratory

「TextGeneration-107403504.ipynb」的副本

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

留言 共用 編輯

+ 程式碼 + 文字

```
for i in range(words):
    next_input = tf.expand_dims(input,axis=0)
    predicts = model(next_input)
    predicts = predicts[:,-1,:]
    predicts /= temperature
    result = tf.random.categorical(
        predicts,num_samples=1
    )
    chinese_ind = tf.squeeze(result).numpy()
    print(index_2_word[chinese_ind],end='')
    input = input+[chinese_ind]
    input = input[-seq_len:]

init_seq = "煙雨濛濛"
init_seq_ind = [word_2_index[w] for w in init_seq]
input = init_seq_ind[-seq_len:]

generateWords(input,500)
```

煙雨濛濛的走了出去。我在客廳裡坐了下來，立即，四週之後，我們才能吐出一口氣。

書桓在我身邊，我轉過頭，看到我回頭。她扭著身子，露出一口煙和惶惑的，又一次又一次。這時，我們的同學，只是我們喪中都是個鬼，可是，我腦子中越來越沉重，幾次越氣越氣，氣不得像一個我的。爸爸，「我想先宣媽，」我說：「我一點都沒有不寧了。」

「你懷疑我——」雪姨大聲的喊。

「那麼，你幫我去就是我，我是個沒有人性的女兒！」

「雪姨！」爸爸興奮的插了進來說：「你想要！我要讓你死！」

「我不是！」他們繼續走了一段，方瑜說：「你是不是需要一個助理？」

「我不做你！」他笑著說：「我也不知道你來了，你要我們一起，給我一個意外！」

「你是嗎？」我有點氣憤：「你認為我是個危嗎？」

「是的，在金錢方面很貧窮，在人母親，」在一個小包車前面，我的手從沙發中抬起來，他的眼睛哀傷而淒苦。我望著他，於是，他又張開眼睛來看看我，低聲的說：「你再多說也沒用，你要就把錢，我不是再在你

「哦，我的天！」我嘆口氣。「那麼，你斷定有個外來的共謀犯，是人、夢萍、夢萍、夢萍，我們再重新認識，重新戀愛多好！」這是

```
import time
while True:
    time.sleep(5)
    pass
```

執行中 (已持續 1 分鐘 46 秒) Cell

## • Stock RNN

colab.research.google.com

HW2 - Google 雲端硬碟 「StockRNN-學號.ipynb」的副本 - Colaboratory 「TextGeneration-107403504.ipynb」的副本 - Colabor... python - expected ndim=3, found ndim=2 - Stack Over...

「StockRNN-學號.ipynb」的副本 ☆

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

留言 共用 設定 幫助

+ 程式碼 + 文字

輸入陣列或標量: (batch,seq\_len,1)

- 輸出維度應為: (batch,seq\_len,1)

```
# todo

model = tf.keras.Sequential()

model.add(tf.keras.layers.LSTM(units = 32, return_sequences = True, input_shape=(seq_len,1)))
model.add(tf.keras.layers.LSTM(units = 64, return_sequences = True))
model.add(tf.keras.layers.LSTM(units = 64, return_sequences = True))

model.add(tf.keras.layers.Dense(len(train_ds)))

model.summary()
```

Model: "sequential\_11"

Layer (type)	Output Shape	Param #
lstm_23 (LSTM)	(None, 15, 32)	4352
lstm_24 (LSTM)	(None, 15, 64)	24832
lstm_25 (LSTM)	(None, 15, 64)	33024
dense_7 (Dense)	(None, 15, 7)	455

Total params: 62,663  
Trainable params: 62,663  
Non-trainable params: 0

0 秒 完成時間: 下午12:07

HW2 - Google 雲端硬碟

「StockRNN-學號.ipynb」的副本 - Colaboratory

「TextGeneration-107403504.ipynb」的副本 - Colabor...

python - expected ndim=3, found ndim=2 - Stack Over...

「StockRNN-學號.ipynb」的副本

★

檔案

編輯

檢視畫面

插入

執行階段

工具

說明

已儲存所有變更

留言

共用

設定

+ 程式碼

+ 文字

RAM

磁碟

編輯

6. 評估模型

```

l1, = plt.plot(history.history['loss'])
l2, = plt.plot(history.history['val_loss'])
plt.legend(handles=[l1,l2],labels=['loss','val_loss'],loc='best')
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')

plt.show()

```

model loss

epoch	loss	val_loss
0.0	0.14	0.06
2.5	0.06	0.02
5.0	0.04	0.015
7.5	0.03	0.01
10.0	0.025	0.008
12.5	0.02	0.007
15.0	0.018	0.006
17.5	0.015	0.005

```

[97] test_company = test_df[test_df["company"] == company_id]
test_company.head()

```

	company	date	price
116	2317	2021/2/1	113.0
115	2317	2021/2/2	116.5

0 秒

完成時間：下午12:07

RNN

7

HW2 - Google 雲端硬碟

「StockRNN-學號.ipynb」的副本 - Colaboratory

「TextGeneration-107403504.ipynb」的副本 - Colabor...

python - expected ndim=3, found ndim=2 - Stack Over...

CO

「StockRNN-學號.ipynb」的副本

檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

留言 共用 編輯

+ 程式碼 + 文字

RAM 磁碟

編輯

### 7. 做預測 (20分)

```
# 預測 Test data
# 現在我們取出了選定公司最後seq_len天
# 我們模型透過讀取seq_len天後預測出seq_len+1天的價格

# 將日期保存在dates中 (list)
# 將真實價格保存在real_prices中 (list)
# 將預測價格保存在pred_prices中 (list) *注: 請填入經過正規化後的價格

# input是一個經過**正規化**的真實價格list, 請試著遍歷test_company來填滿上述三個list
# 可以TextGeneration的generateWords function, 概念一樣, 但是這次input請全部都填入"真實價格"

init_price = company.iloc[-seq_len:, :]["price"].values
dates = []
real_prices = []
pred_prices = []
input = scalar.transform(init_price)

for index, row in test_company.iterrows():
    # 我已經幫你把real_price和date放入list
    real_price = row["price"]
    real_prices.append(real_price)
    date = row["date"]
    dates.append(date)
    next_input = tf.expand_dims(input, axis=-1)
    next_input = tf.expand_dims(next_input, axis=0)
    predicts = model(next_input)
    predicts = predicts[0, -1, 0].numpy()
    pred_prices.append(predicts)
    input = input + [scalar.transform_one(real_price)]
    input = input[-seq_len:]

# 參考步驟:
# 1. 把input轉維度成一個新變數next_input丟入model
# 2. 獲得的pred取出seq中最後一個時間點的output數值 (提示: [0 -1 0])
```

0秒 完成時間: 下午12:07



+ 程式碼 + 文字

RAM 磁碟 編輯

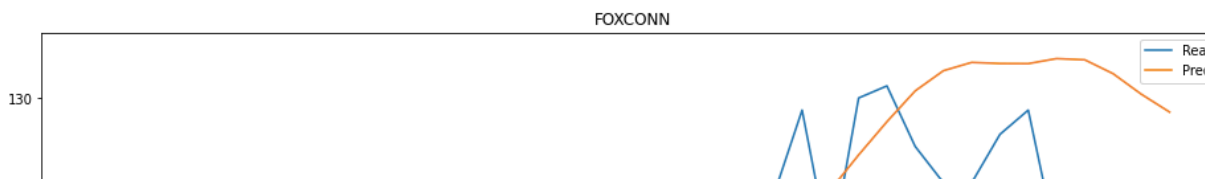
```
print(f"dates : {dates[:10]}")
print(f"real_prices : {real_prices[:10]}")
print(f"pred_prices : {pred_prices[:10]}")
```

```
dates : ['2021/2/1', '2021/2/2', '2021/2/3', '2021/2/4', '2021/2/5', '2021/2/17', '2021/2/18', '2021/2/19', '2021/2/22', '2021/2/23']
real_prices : [113.0, 116.5, 116.5, 114.0, 113.5, 114.0, 113.0, 110.0, 111.0, 111.5]
pred_prices : [1.0156329, 0.9917656, 0.97106713, 0.95806956, 0.94723386, 0.94020146, 0.9367281, 0.93328726, 0.9250408, 0.91286314]
```

#### • 預期結果:

```
dates : ['2021/2/1', '2021/2/2', '2021/2/3', '2021/2/4', '2021/2/5', '2021/2/17', '2021/2/18', '2021/2/19', '2021/2/22', '2021/2/23']
real_prices : [905.0, 931.0, 950.0, 946.0, 950.0, 980.0, 978.0, 964.0, 951.0, 965.0]
pred_prices : [0.9349597, 0.9338685, 0.94261205, 0.9576455, 0.9740759, 0.987802, 1.0038555, 1.0165175, 1.0234691, 1.0223937]
```

```
plt.figure(figsize=(16, 8))
l1, = plt.plot(dates, real_prices)
l2, = plt.plot(dates, scalar.recover(pred_prices))
plt.xticks(list(include_last_iter(0, len(dates), 3)), rotation=30)
plt.legend(handles=[l1, l2], labels=['Real', 'Pred'], loc='best')
plt.title(companys[str(company_id)])
plt.xlabel("Date")
plt.ylabel("price")
plt.show()
```



0 秒 完成時間：下午12:07



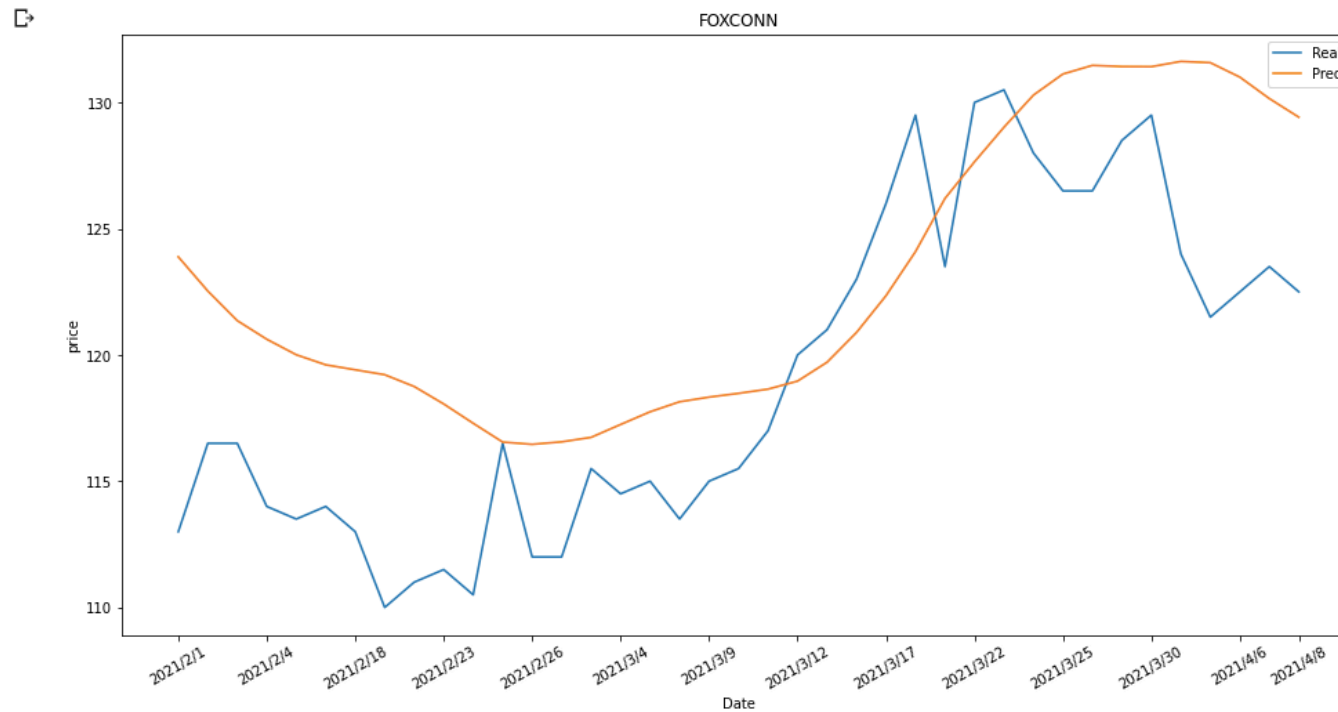
「StockRNN-學號.ipynb」的副本 ☆  
檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更

留言 共用 編輯

+ 程式碼 + 文字

✓ RAM 磁碟 編輯

```
plt.ylabel("price")  
plt.show()
```



- 預期成果

照理來說不會100%猜中，猜中幾乎100%可能是不小心偷看答案了

MEDIATEK

✓ 0 秒 完成時間：下午12:07

