

**以 Gated Recurrent Unit (GRU) 應用於
美國那斯達克指數及投資人情緒指標預測台灣加權指數**

**Apply Gated Recurrent Unit to Predict
Taiwan Se Weighted Index (TWSE) by Nasdaq Composite Index
(NASDAQ) and Benchmarks of Investors' Sentiments**

東吳大學 財務工程與精算數學系

指導教授：林忠機 (Chung-Gee Lin)

研究生：許沛萱 (Pei-Hsuan Hsu)

民國 108 年 6 月 17 日

目錄

壹、 緒論	4
一、 研究動機	4
二、 目的	4
貳、 文獻回顧	5
參、 研究方法	5
一、 研究設計	5
二、 資料來源	6
三、 Gated Recurrent Unit(GRU)介紹	6
肆、 實證分析	7
一、 資料預處理(Data Preprocessing)	7
二、 建立模型與預測結果	8
伍、 結論與建議	10
陸、 文獻參考	10

圖目錄

FIGURE 1 GRU CELL	6
FIGURE 2 模型誤差	8
FIGURE 3 模型誤差	8
FIGURE 4 股價預測	9

表目錄

TABLE 1 加權指數成分股市值前 20 名	4
TABLE 2 變數整理	7

摘要

在全球化的衝擊下，各國之間的經濟狀況息息相關，因此了解國際經濟情勢對於現代投資人而言是非常重要的。而台灣經濟時常受到美國經濟的變動影響，因此本研究將使用循環類神經網路(Recurrent Neural Network, RNN) 改良後的 Gated Recurrent Unit (GRU) (Cho et al., 2014)，並使用西元 2007 年至 2018 年美國那斯達克指數，以及台灣投資人的情緒指標預測台灣加權指數的趨勢。

關鍵字：循環類神經網路、GRU、美國那斯達克指數

Abstract

Under the impact of globalization, there is a close relationship between economies. Therefore, comprehending international economic circumstances is important to the modern investors. Furthermore, Taiwan's economy is often affected by American's, so this research is going to predict Taiwan Se Weighted Index (TWSE) with benchmarks of investors' sentiments and NASDAQ from 2006 to 2018 by using Gated Recurrent Unit (GRU) (Cho et al., 2014), improved Recurrent Neural Network(RNN).

Key Words: Recurrent Neural Network, Gated Recurrent Unit, NASDAQ

壹、緒論

一、研究動機

現代各國間的經濟均息息相關，以台灣股市為例，當美國有新的經濟政策或美國股市出現巨大變動，台灣股市通常會明顯地受到影響。以美國聯準會(The Federal Reserve System)升息對台灣股市的影響為例，台灣股市大部分的資金來源為外資，當美國實施升息時，外資將會減少在外流動的資金，因此台灣股市資金減少，進而受到影響。

二、目的

台灣加權指數的權值以電子業為主，從 Table 1 可以發現台積電占的比例極高，對大盤的走勢影響大且台積電與美國多家高科技公司皆有貿易關係，此外那斯達克綜合指數又有許多高科技公司上市，因此本研究將利用循環類神經網路以美國那斯達克綜合指數及台灣投資人情緒指標進行預測台灣加權指數。藉此研究那斯達克指數和台灣加權指數的關係。

加權指數成分股與市值比重							
排行	代號	名稱	市值佔大盤比重	排行	代號	名稱	市值佔大盤比重
1	2330	台積電	21.19%	11	1216	統一	1.46%
2	2317	鴻海	4.21%	12	2891	中信金	1.42%
3	6505	台塑化	3.77%	13	2002	中鋼	1.34%
4	2412	中華電	2.95%	14	3045	台灣大	1.33%
5	1301	台塑	2.26%	15	2454	聯發科	1.32%
6	1326	台化	2.19%	16	2886	兆豐金	1.25%
7	2882	國泰金	2.18%	17	2912	統一超	1.20%
8	1303	南亞	2.10%	18	2308	台達電	1.19%
9	2881	富邦金	1.78%	19	3711	日月光投控	0.96%
10	3008	大立光	1.68%	20	2892	第一金	0.89%

資料來源：<https://histock.tw/stock/taixproportion.aspx>

Table 1 加權指數成分股市值前 20 名

貳、文獻回顧

一、美國股市與台灣股市之關聯性

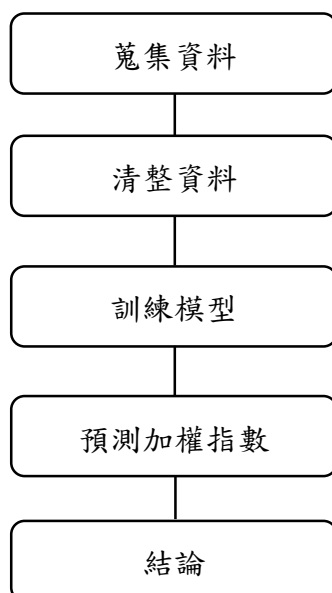
李敏生(2000) 研究 NASDAQ 綜合指數、NASDAQ 電腦類指數及 NASDAQ-100 指數對於台灣上市、上櫃大盤以及電子類指數報酬率與波動性的影響，發現 NASDAQ 綜合指數報酬率對於台灣股市有顯著的正相關。

二、Gated Recurrent Unit

Yumo Xu & Shay B. Cohen(2018) 利用自然語言分析社群網站推特(Tweets)的文字和股價資料來做股價預測，結果顯示使用 Gated Recurrent Unit 可以顯著的提升訓練模型的效率。

參、研究方法

一、研究設計



二、資料來源

本研究之台灣加權指數及台指 VIX 資料取自台灣經濟新報 TEJ 資料庫，那斯達克綜合指數資料取自雅虎財經 Yahoo Finance，訓練集資料期間為 2007 年 1 月 1 日至 2018 年 12 月 31 日，測試集資料期間為 2019 年 1 月 1 日至 2019 年 4 月 30 日。

三、Gated Recurrent Unit(GRU)介紹

GRU 為遞迴類神經網路(Recurrent Neural Network, RNN) 的一種，是長短期記憶(Long Short-Term Memory, LSTM)的簡化版，利用 update gate 來代替 LSTM 中的 forget gate 和 input gate。

其結構如下：

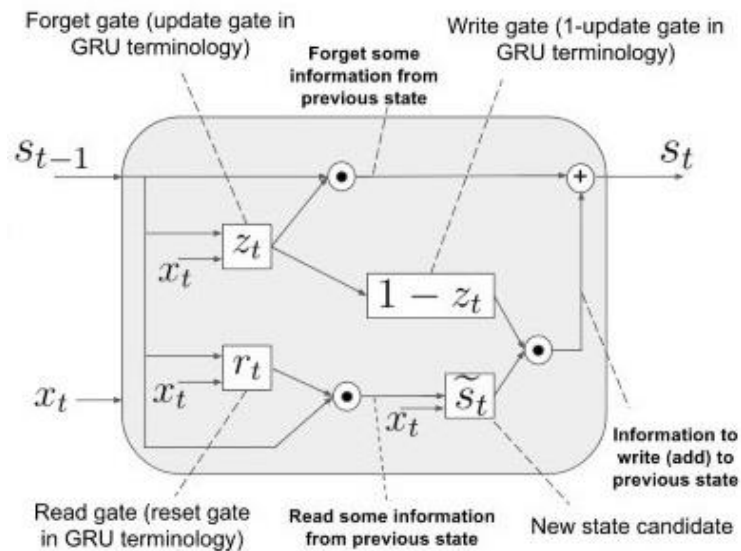


Figure 1 GRU Cell (資料來源：<https://is.gd/VvtAZI>)

GRU 之流程如下：

- 一、 首先輸入前一期的結果 S_{t-1} 和當期的資訊 x_t ，再經過兩個 gate， r_t 為 reset gate， z_t 為 update gate，公式和 LSTM 一樣：

$$r_t = \sigma(W_r S_{t-1} + U_r x_t + b_r)$$

$$z_t = \sigma(W_z S_{t-1} + U_z x_t + b_z)$$

其中， σ 為 sigmoid 函數， W 、 U 為權重， b 為 bias

二、 接著計算候選隱藏層(Candidate Hidden Layer) \tilde{S}_t 其公式如下：

$$\tilde{S}_t = \tanh(W(r_t \circ S_{t-1}) + Ux_t + b)$$

\tilde{S}_t 可看成當下的新訊息，並利用 r_t 來控制前一期的資訊 S_{t-1} 要保留多少。

三、 最後利用 z_t 控制保留多少前期隱藏層 \tilde{S}_{t-1} 的資訊：

$$S_t = z_t \circ S_{t-1} + (1 - z_t) \circ \tilde{S}_t$$

S_t 即為當期輸出的資訊。

肆、實證分析

一、 資料預處理(Data Preprocessing)

1. 蒐集資料(2007/1/1~2019/4/30)本研究使用的資料如下表：

TWII、NASDAQ	VIX
調整後開盤價	開盤價
調整後最高價	最高價
調整後最低價	最低價
調整後收盤價	收盤價
指數變動	收盤價變動
成交量	時間描述
成交值	年份
報酬率	月份
報酬率變動	日期
成交量變動	星期
N日標準差	距離資料 第一天日數
N日偏態係數	
N日標準差變動	* N=5, 20, 60, 120
N日偏態係數變動	

Table 2 變數整理

其中，台灣加權指數之調整後收盤價為目標變數，其餘皆為特徵變數。

2. 區分訓練集及測試集資料

訓練集資料期間：2007/1/1~2018/12/31

測試集資料期間：2019/1/1~2019/4/30

3. 特徵正規化(Normalization)

利用 MinMax 函數將特徵值按比例縮放至(0,1)之間，其公式如下：

$$m_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

其中 x_i 為該特徵第 i 個值， m_i 為正規化後的數值

二、 建立模型與預測結果

將清整後的資料丟入模型訓練後，我們得到模型的訓練資料及驗證資料的誤差：

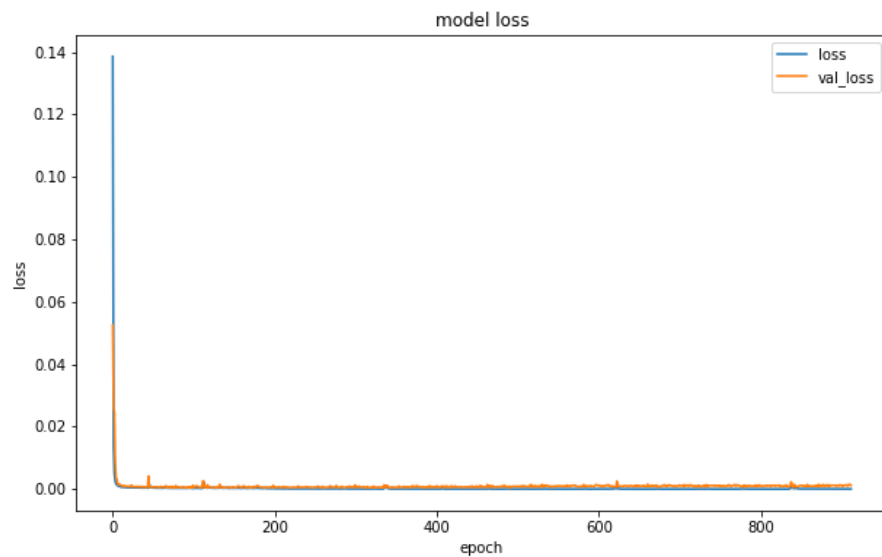


Figure 2 模型誤差

Figure 2 為所有資料的誤差，由於後來的誤差都很小，因此我們抓出第 10 筆以後的數據畫出誤差的圖形，如 Figure 3：

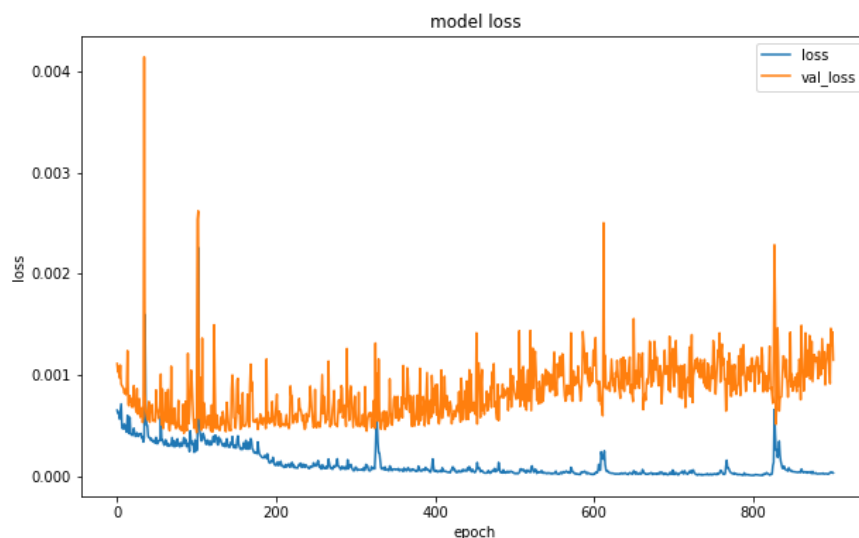


Figure 3 模型誤差(第 10 筆資料以後)

由 Figure 3 可以觀察到訓練資料的誤差為下降的趨勢，而驗證資料的誤差卻有些微的上升，所以有過擬合(Over fitting)的疑慮，但是誤差變動的範圍僅在 0.001 之間，因此本研究忽略模型過擬合的疑慮。最終模型的誤差為 0.12%，將訓練好的模型做預測，得到 Figure 4：

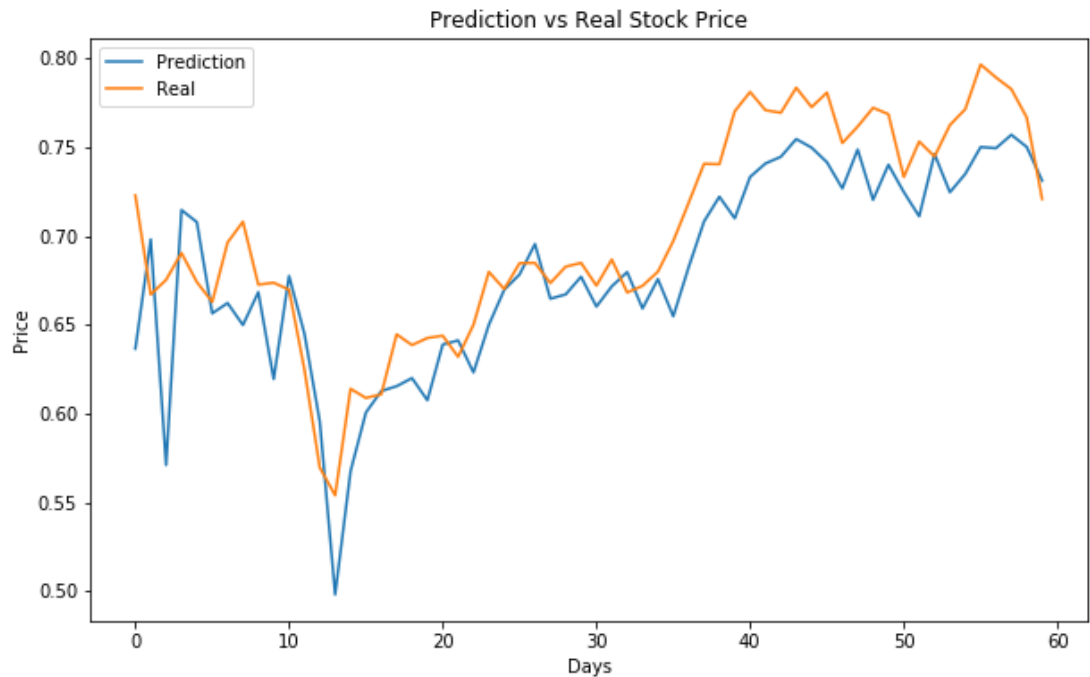


Figure 4 股價預測

由 Figure 4 可以看到未還原股價的預測結果，10 天以前的預測效果並不是很好，誤差極大，但是 10 天之後大致可以抓到股價波動的趨勢，因此此預測模型只能判斷長期大概的趨勢，無法精確的預測隔天漲幅。

伍、結論與建議

從模型的預測結果可以發現利用美國那斯達克綜合指數、台指 VIX 以及台灣加權指數來訓練模型，並未非常精準地捕捉到隔日的漲跌幅，而本研究推論原因有二，其一，研究時所選取的特徵並不能完全描述目標變數和時間的關係，因此無法完美的預測股價；另一個原因可能為模型設定的問題，因時間的限制，並沒有做完整的調參(tuning)，導致此模型的預測效果差。

由於上述原因，後續可以透過調整參數或是增加其他可描述時間的變數來找到最佳的模型，調參的部分像是每層的激活函數該用哪種函數較適合？每層的輸入及輸出神經元數為多少？整個類神經網絡要多少層？模型的損失函數是什麼？還有許多問題都是值得探討的，因此調參為非常重要的步驟，但也是非常耗時的一項工程。

陸、文獻參考

1. 李敏生 (2000)，「NASDAQ 股市對於台灣股市報酬率與波動性的影響」，國立交通大學經營管理研究所碩士論文。
2. Xu, Yumo, and Shay B. Cohen. "Stock Movement Prediction from Tweets and Historical Prices." Melbourne, Australia, jul 2018.