

應用 LSTM 於投資人情緒指標

預測台灣 50 股價及建立交易策略

**Apply Long Short-Term Memory In Predicting Price of
Yuanta Taiwan Top 50 ETF with Benchmarks of
Investors' Sentiment and Developing Trading Strategy**

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 18 日

目錄

壹、	緒論	4
一、	研究動機	4
二、	目的	4
貳、	文獻回顧	4
參、	研究方法	5
一、	研究設計	5
二、	資料來源	5
三、	長短期記憶介紹	6
肆、	實證分析	9
一、	資料預處理(Data Preprocessing)	9
二、	建立模型	10
三、	建立交易時點	12
伍、	結論與建議	13
陸、	文獻參考	13

圖目錄

FIGURE 1: RNN CELL (資料來源： HTTPS://TINYURL.COM/YVVMR9YE)	6
FIGURE 2: LSTM CELL (資料來源： HTTPS://TINYURL.COM/Y4PORBWE)	6
FIGURE 3: LSTM 流程圖一 (資料來源： HTTPS://TINYURL.COM/Q6DCYBC)	7
FIGURE 4: LSTM 流程圖二 (資料來源： HTTPS://TINYURL.COM/Q6DCYBC)	7
FIGURE 5: LSTM 流程圖三 (資料來源： HTTPS://TINYURL.COM/Q6DCYBC)	8
FIGURE 6: LSTM 流程圖四 (資料來源： HTTPS://TINYURL.COM/Q6DCYBC)	8
FIGURE 7 模型誤差	10
FIGURE 8 模型誤差(第 10 筆資料以後)	10
FIGURE 9: LSTM 預測股價結果	11
FIGURE 10: 預測股價與真實股價比較	12

表目錄

TABLE 1: 變數整理	9
TABLE 2: 多空判斷	12
TABLE 3: 交易結果整理	12

摘要

近年來，由於許多人認為投資人並非理性，必須由心理學的角度來探討投資人的行為，因此行為財務學漸漸興起，因此本研究將結合行為財務學及機器學習技術，使用長短期記憶(Long Short-Term Memory, LSTM)以 2006 年至 2018 年投資人情緒指標 - 歷史偏態及峰態系數 - 及價量資料來預測 2019 年 1 月至 4 月台灣 50 股價，進而做投資策略來探討 LSTM 應用於股價預測之績效。

關鍵字：行為財務學、長短期記憶

Abstract

Recently, due to the emerging of behavioral finance and techniques of machine learning, this research is going to combine these two fields to predict Yuanta Taiwan Top 50 ETF. The examination will apply LSTM to predict the price of ETF from January to April in 2019 by using benchmark of investors' sentiments from 2006 to 2018. Besides, we're going to build a trading strategy to assess the performance of LSTM.

Key Words: Behavioral Finance, LSTM

壹、緒論

一、研究動機

由於機器學習的快速發展，有越來越多的模型被發展出來，近年在金融領域的研究更是經常使用類神經網路來做股價預測，其中，像是遞迴類神經網路(Recurrent Neural Network, RNN)便是設計用來處理時間序列的神經網路。因此本研究將嘗試使用由 RNN 發展出的長短期記憶(Long Short-Term Memory, LSTM)應用於台灣股價的預測。

二、目的

在搜尋利用機器學習來預測股價的相關資料時，發現許多模型放入的特徵皆只有一般的價量資料，因此本研究將加入具有描述投資人情緒的指標——偏態係數——來訓練模型，並觀察是否可提升預測的準確度。本次所使用的標的為元大台灣 50 ETF，希望可以利用 2006 年 1 月 1 日到 2018 年 12 月 31 日的資料訓練模型，並預測 2019 年 1 月 1 日到 4 月 30 日的股價。

貳、文獻回顧

一、長短期記憶於金融之應用

Qun Zhuge, Lingyu Xu and Gaowei Zhang (2017) 利用上海證券交易所綜合股價指數以及情緒指標資料，當作 LSTM 模型的輸入資料，並成功的預測股價開盤價。

參、研究方法

一、研究設計



二、資料來源

本研究之元大台灣 50 股價資料取自台灣經濟新報 TEJ 資料庫，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、長短期記憶介紹

長短期記憶(LSTM)為遞迴類神經網路(Recurrent Neural Network, RNN) 的一種，其中，RNN 的結構如下：

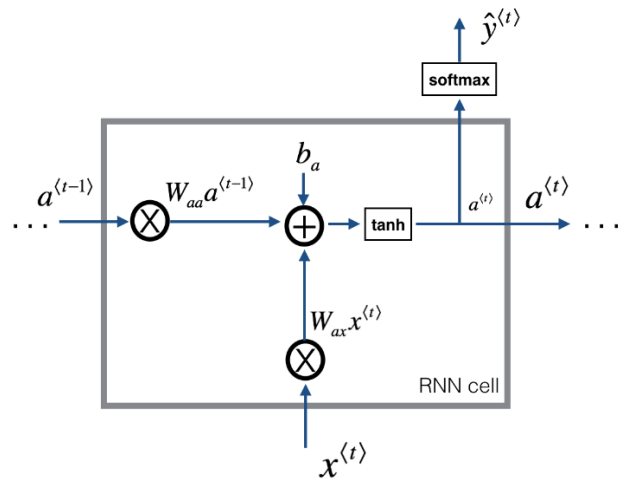


Figure 1: RNN Cell (資料來源：<https://tinyurl.com/yyvmr9ye>)

LSTM 的結構如下：

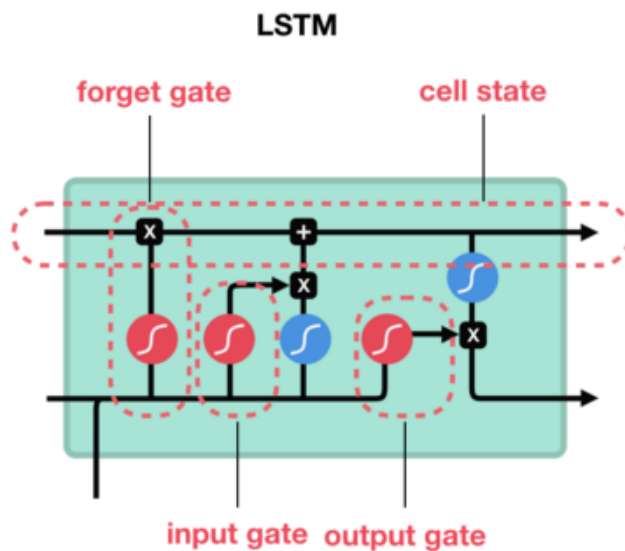


Figure 2: LSTM Cell (資料來源：<https://tinyurl.com/y4porbwe>)

由 Figure 1 和 Figure 2 可看出 LSTM 的神經元比原始的 RNN 多了 Input Gate、Forget Gate、Cell State 和 Output Gate，而 LSTM 的流程如下所示：

一、由 Forget Gate 決定要保留多少以前記憶的資訊 C_{t-1}

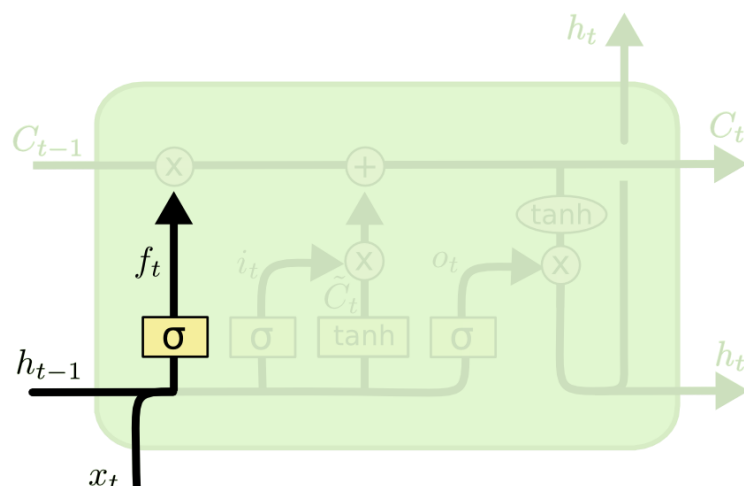


Figure 3: LSTM 流程圖一 (資料來源：<https://tinyurl.com/q6dcybc>)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

其中， σ 為 Sigmoid 函數

二、產生當時記憶的候選值(Memory Cell Candidate) \tilde{C}_t ，並用 Input Gate i_t 決定哪些資訊需要保留

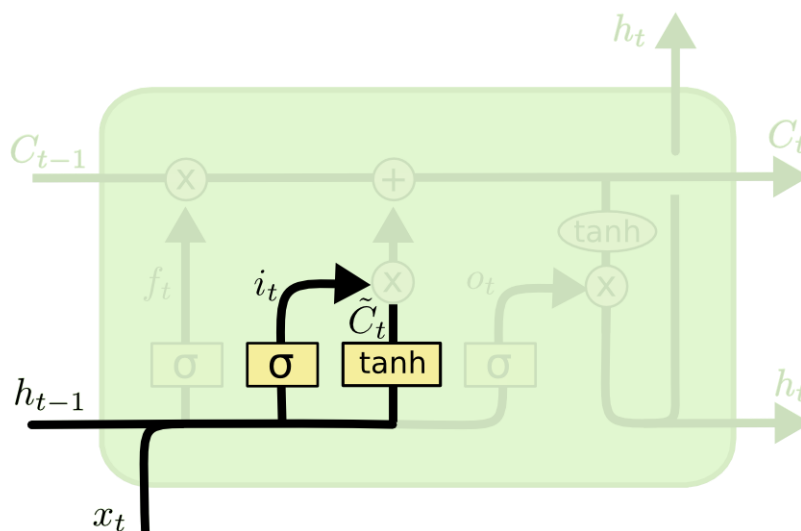


Figure 4: LSTM 流程圖二 (資料來源：<https://tinyurl.com/q6dcybc>)

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

三、產生當期記憶資訊 C_t

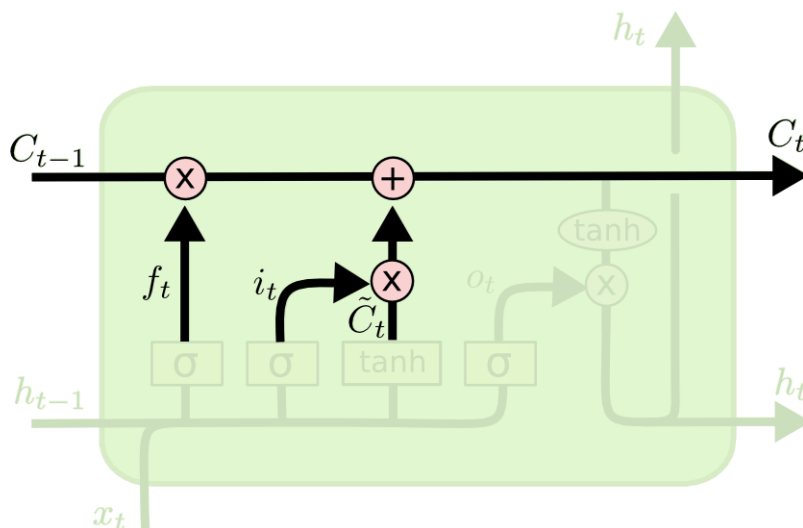


Figure 5: LSTM 流程圖三 (資料來源：<https://tinyurl.com/q6dcybc>)

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

四、利用 Output Gate o_t 決定是否將當期資訊加入長期記憶中

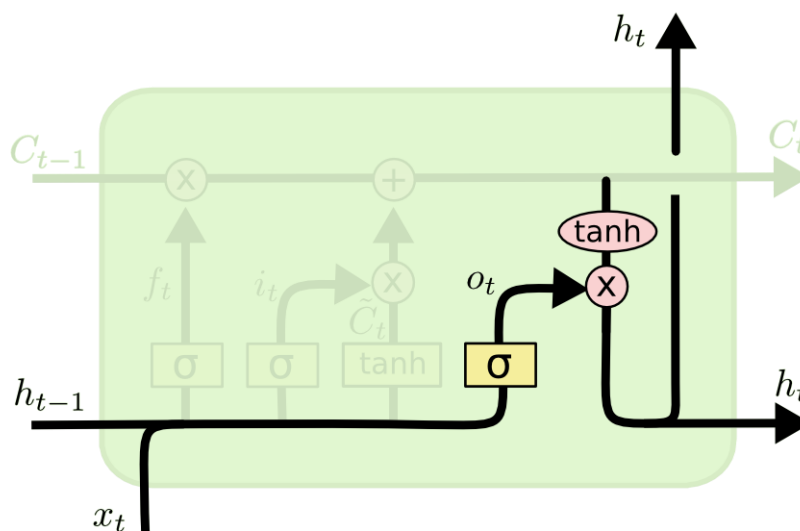


Figure 6: LSTM 流程圖四 (資料來源：<https://tinyurl.com/q6dcybc>)

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

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：

0050 ETF	時間描述
調整後開盤價	年份
調整後最高價	月份
調整後最低價	日期
調整後收盤價	星期
報酬率	距離資料 第一天日數
報酬率變動	
N日標準差	* N=5, 20, 60, 120
N日偏態係數	
N日標準差變動	
N日偏態係數變動	

Table 1: 變數整理

其中，台灣元大 50 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

將清楚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：

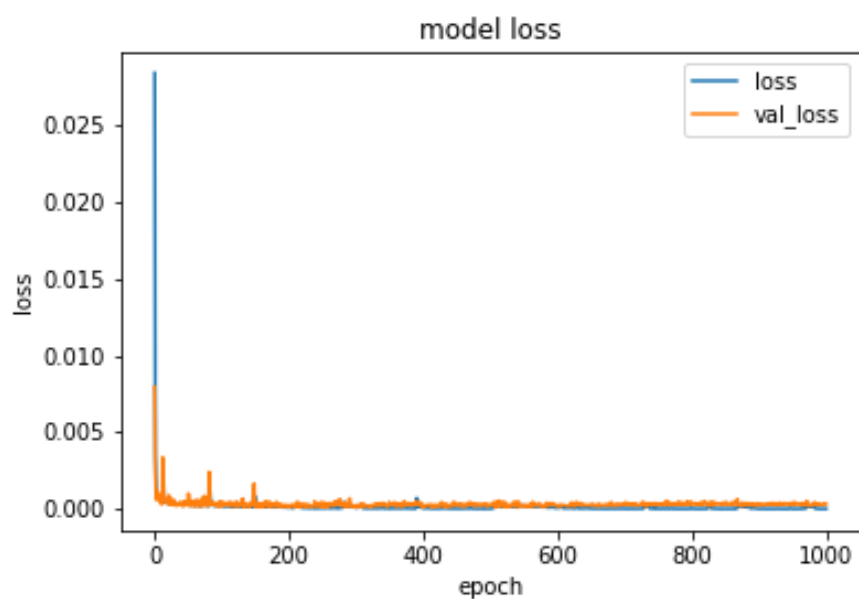


Figure 7 模型誤差

Figure 7 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 8 所示：

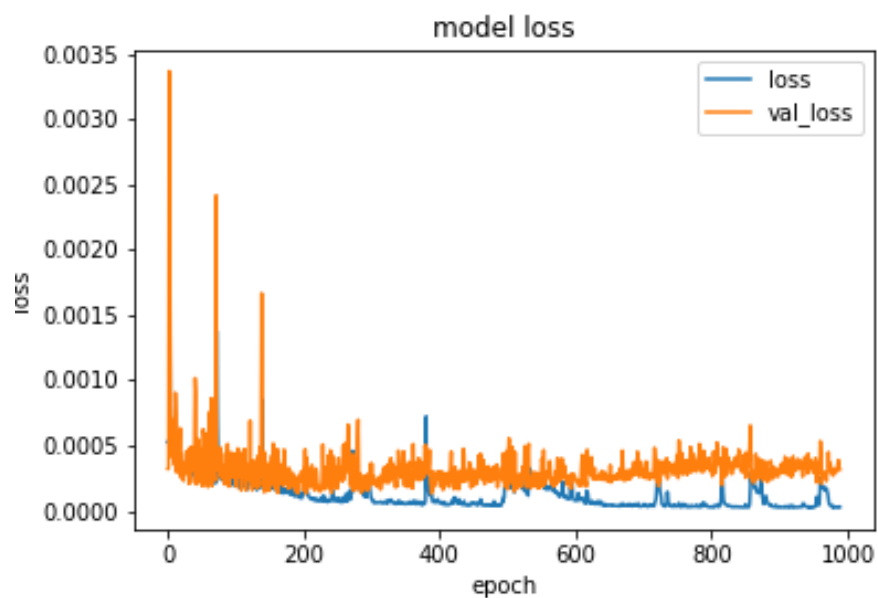


Figure 8 模型誤差(第 10 筆資料以後)

由 Figure 8 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 0.0005 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 0.3%，平均平方誤差(Mean Square Error)為 0.152，將訓練好的模型做預測，得到 Figure 9：

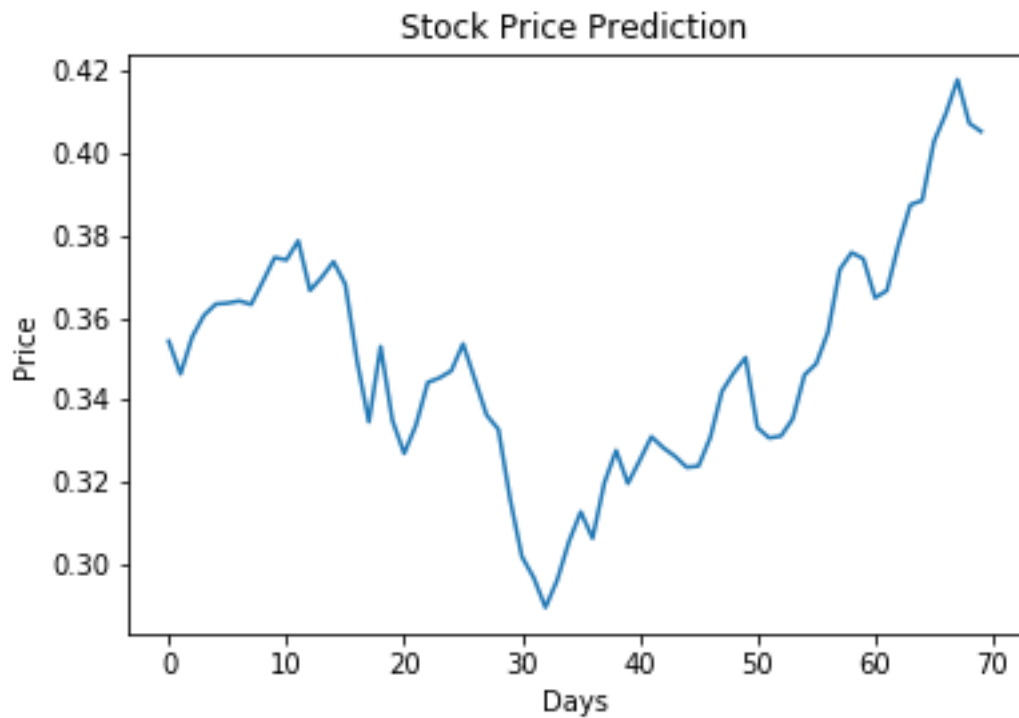


Figure 9: LSTM 預測股價結果

三、建立交易時點

根據 Figure9 的結果，本研究將預測期間一多空趨勢判斷，如下表所示：

期間	趨勢
第0到第15天	多頭
第15天至第30天	空頭
第30天至第70天	多頭

Table 2: 多空判斷

本研究測試的交易時點為多頭時期在預測時低點買進，預測之糕點賣出，因此僅在預測之 0~15 天及 30~70 天交易，而交易時點和結果如下：

交易	買進日期	買進收盤價	賣出日期	賣出收盤價	利潤*
第一次	2019/1/9	72.75	2019/1/23	73.7	0.95
第二次	2019/3/6	76.95	2019/4/26	82.45	5.5

*利潤未考慮交易成本

Table 3: 交易結果整理

從 Table3 可發現利用預測結果來做交易時點買賣可以得到正報酬，以下再來看真實股價趨勢與預測結果做比較：

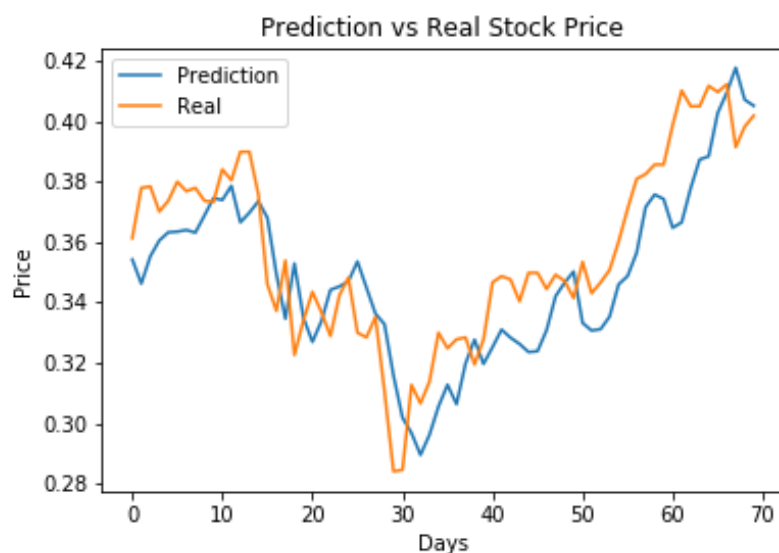


Figure 10: 預測股價與真實股價比較

從 Figure10 可以發現，LSTM 模型可抓到大概的趨勢，但是無法捕捉到細微的波動，因此可以用來做多空趨勢的判斷。

伍、結論與建議

從模型的預測結果可以發現 LSTM 模型可以預測到未來的多空趨勢，但是若要做到精準的股價預測還需要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。或是可以改為和 Qun Zhuge et al.(2017) 一樣預測開盤價，或許可得到更佳的结果。

陸、文獻參考

1. Qun Zhuge, Lingyu Xu and Gaowei Zhang (2017), “LSTM Neural Network with Emotional Analysis for Prediction of Stock Price”, <https://pdfs.semanticscholar.org/1a08/bcc2e0866e64d7b4d8d7a4c670faa26ab13d.pdf>, 2017