

Predict Stock-皮弟是他 機器學習暨各國股市指標之個股預測

專題編號：107-CSIE-S019

執行期限：106年第1學期至107年第1學期

指導教授：王正豪 教授

專題參與人員： 104590027 陳志芳

104590030 蔡銘祐

104590047 林軒宏

一、摘要

本專題主要是希望求出各國的股票指數歷史資料與臺灣上市的股票之間的關聯性，並且加入歷史資料，再將此資料(關聯係數+各國指數歷史資料)，運用機器學習的全鏈結層技術，訓練出臺灣上市各股的預測模型，並再每日臺股開市(9:00 am)之前，提供每日的各國股市資訊給預測模型，預測今日收市(1:30 pm)之前的漲跌幅度。

關鍵詞：機器學習、國際股市、個股預測

二、緣由與目的

現今的股票市場，大多以人為操作為主，但人為操作，往往需要觀察幾周到幾個月的時間，甚至更長的時間才能找出自己的操盤策略，再加上臺灣的股票市場非常容易受到國際股市的影響，且各股股票價格的漲跌受籌碼面(人為操作)的影響，遠比基本面(公司營收)高出許多。

三、使用技術

(一) PyMySQL：

Python 中的套件，用於將爬蟲爬下來國際股市資料從資料庫中提取出來；模型建立好之後，將關聯性係數存入資料庫；每日跑完預測模型之後，將預測的漲跌幅存入資料庫。

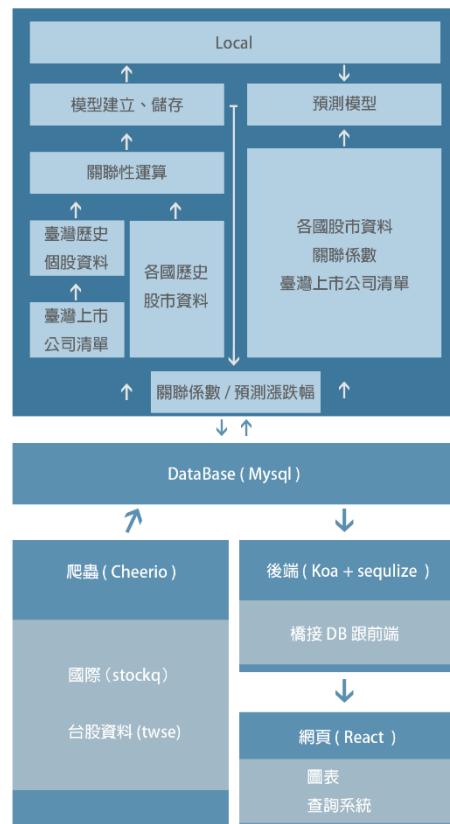
(二) Keras：

Python 中的套件，用於做機器學習及類神經網路的運用，在本專題利用 Keras 實作全連結神經網路進行資料訓練、模型建立及漲跌幅預測。

(三) Node.js：

利用 Node.js 發 Request 讀取網頁的統計資料，並透過 Cheerio(解析網頁標籤)做網頁標籤的擷取，之後將想要的網頁資料，經資料處理後，匯入資料庫。

四、架構流程



由於本專題在資料溝通的部分，揭示透過資料庫傳遞，因此，在程式架構上大致分為三個區域，分別是「模型及演算法」、「資料庫」及「爬蟲及網頁前端」：

(一) 模型及演算法：

1. 模型建立：

透過從資料庫取得臺灣上市公司清單及各國股市歷史資料，並利用清單再到資料庫中取得目前各上市公司的股票資訊；將取得的各上市股票與各國歷史資料做相關係數運算，並將相關係數及資料丟入神經網路中訓練，模型訓練完成後，將臺灣各股相對於各國資料的相關係數儲存回資料庫，並將模型儲存於本機端。

2. 每日預測：

在每個工作日早上九點開市前，提取資料庫中各國股市資料，並用相對應的預測模型做預測，並將結果存入資料庫中。

(二) 資料庫：

存有「預測資料儲存」、「各股相關係數」、「臺灣上市公司清單」、「各國歷史資料」及「臺灣股票歷史資訊」。

(三) 爬蟲及網頁前端：

1. 歷史資料：

將往年的各國指標及台灣股市的歷史資料匯入資料庫。

2. 每日更新：

每個工作日早上九點開市前，將各國股市資訊匯入資料庫。

3. 網頁前端：

利用 React.js 統整 Koa.js 傳送的資料庫資訊、Node.js 傳送的各個股票相關的新聞以及利用 React.js 中的套件將所得到的資訊做顯示。

五、實際畫面



圖1 實際網頁畫面

六、結論

雖然現在已經能穩定的運作所有的功能，但是由於所能利用的股市資訊實在太少，造成必須將手上所擁有的所有資訊都當作訓練集，除此之外，也因為如此而造成模型 overfitting 十分的嚴重，因此，未來希望能用利用其他的方式改善，如：CNN、RNN 以及資料量化等方式，改善模型，也會試著加入其他不同的因素，來增加資料的連結性，以做出更加客觀預測的模型。

七、參考文獻

[1] Keras

<https://keras.io/>

[2] PyMySQL

<https://github.com/PyMySQL/PyMySQL>

[3] Pearson Correlation Coefficient

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

[4] Fully Connected Layer

https://leonardoaraujosantos.gitbooks.io/official-intelligence/content/fc_layer.html