

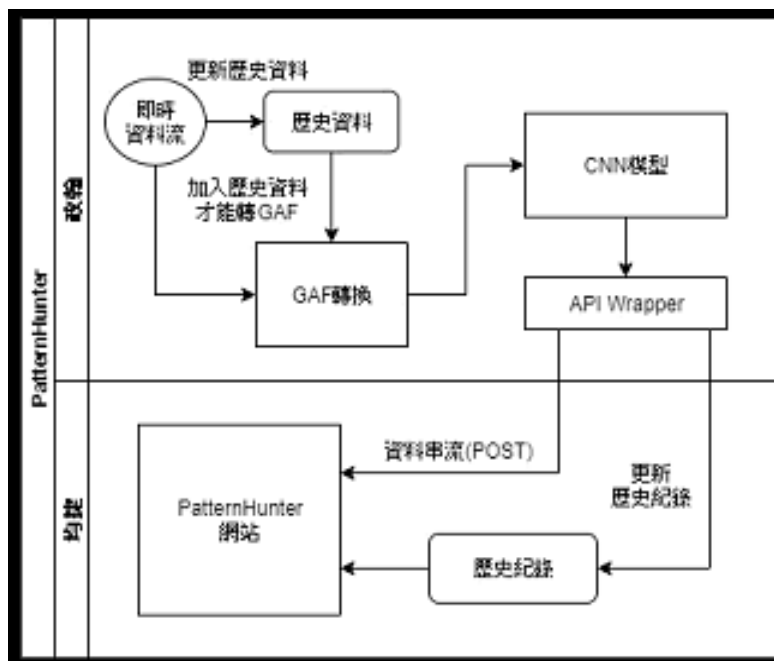
Pattern Hunter

一、研究說明

期望開發出一套交易訊號偵測系統給予投資人使用，該系統串流了某個時間尺度的標的資產（可為分鐘、小時、日或周資料），並使用深度學習影像辨識的方式，按該時間尺度的頻率偵測標的資產在形態學上交易訊號，如在某區段偵測出符合形態學上交易訊號，系統則會將該訊號提供給投資人作為買賣判斷的依據。該研究的發想在於形態學交易大多需要仰賴人為主觀的判斷，尚未有一套技術指標能夠精準地量化，光靠人為的判斷難免會失真而且費時費力，因此如果能透過深度學習的方法來辨識形態學的影響，投資人不但能增加投資判斷的精準度，也能更有效率地進行形態學交易。

二、系統流程

本系統分成前端與後端兩項開發。測試的標的資產為 S&P 500，時間尺度為日資料。前端的部分，首先透過 financialmodelingprep 所提供的 API 載入標的資產的開盤、最高、最低、收盤價資料 (ohlc)；接著將 ohlc 做 GAF 轉換，將一維的時間序列資料轉換為二維的圖像資料；接著透過 CNN 模型進行訓練；隨後再透過 financialmodelingprep 的 API 串流即時資料，模型將會依串流的即時資料進行形態學交易訊號預測。後端的部分主要進行網站開發以及使用者介面優化。本篇研究只包含前端開發的部分。



2.1 歷史資料蒐集

使用 financialmodelingprep 所提供的 API，資產標的為 S&P500，時間尺度以「日」為單位，時間範圍從 2015/2/18 到 2020/2/13，總共 1258 筆。其中載入的資料時間為美國時間，為了提供台灣投資人於本地使用，已將美國時區時間轉台灣時區時間 UTC+8，而夏令時則再將台灣時區的開盤時間往前調整一小時。

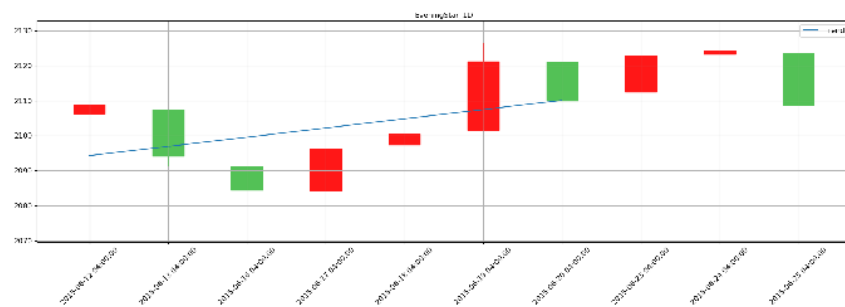
美國股市的交易時間於冬令時，美東時間從早上 9:30 到下午 4:00，換算成台北時間為晚上 10:30 到早上 5:00；於夏令時，美東時間從早上 9:30 到下午 4:00，換算成台北時間為晚上 9:30 到早上 4:00。另外美國的夏至期間是從 03 月 10 日 02:00 星期日上午開始 (美國時間 3 月 11 日交易時間才調整)到 11 月 03 日 02:00 星期日上午結束。

| | date | open | high | low | close | volume | realbody |
|---|---------------------|-------------|-------------|-------------|-------------|--------------|-----------|
| 0 | 2015-02-18 05:00:00 | 2096.469971 | 2101.300049 | 2089.800049 | 2100.340088 | 3.361750e+09 | 3.870117 |
| 1 | 2015-02-19 05:00:00 | 2099.159912 | 2100.229980 | 2092.149902 | 2099.679932 | 3.370020e+09 | 0.520020 |
| 2 | 2015-02-20 05:00:00 | 2099.250000 | 2102.129883 | 2090.790039 | 2097.449951 | 3.247100e+09 | -1.800049 |
| 3 | 2015-02-21 05:00:00 | 2097.649902 | 2110.610107 | 2085.439941 | 2110.300049 | 3.281600e+09 | 12.650147 |
| 4 | 2015-02-24 05:00:00 | 2109.830078 | 2110.050049 | 2103.000000 | 2109.659912 | 3.093680e+09 | -0.170166 |

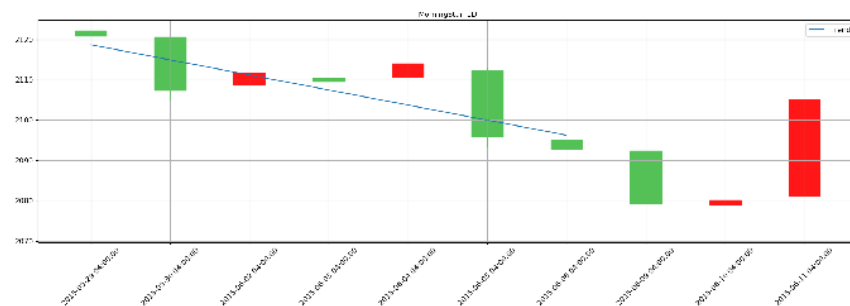
2.2 形態學訊號偵測

偵測範圍為每 10 根 K 棒為單位 (十天的 ohlc)，偵測的形態包含：EveningStar、MorningStar、BearishHarami 與 BullishHarami。形態學的定義是參考 Stephen W. Bigalow 所著的 The Major Candlesticks Signal 該本書。

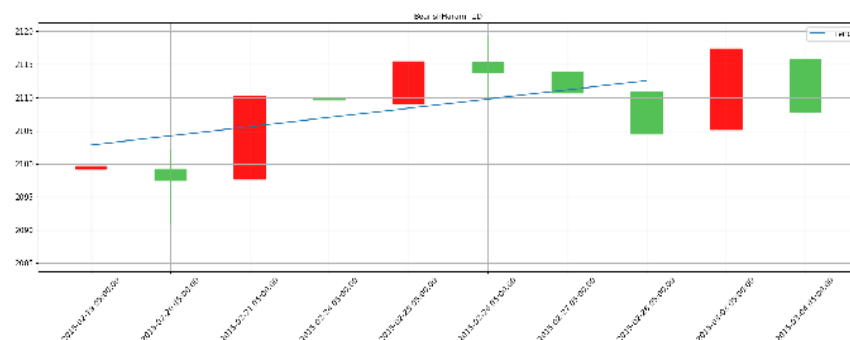
EveningStar 需符合以下要件：1. 趨勢上漲、第一根上漲、第二根上漲、第三根下跌 2. 第二根開盤價 > 第一根 1/2 的價格 3. 第三根收盤價 < 第一根 4/5 的價格。



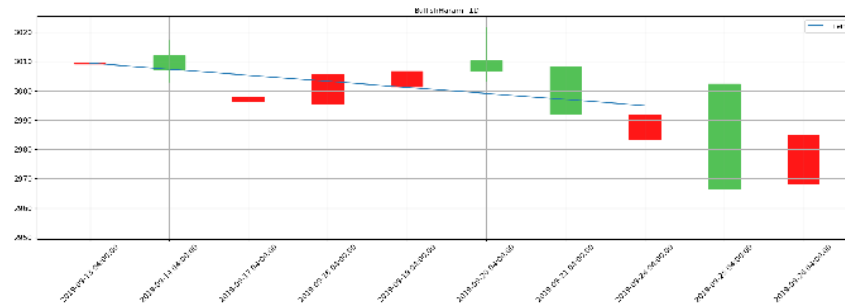
MorningStar 需符合以下要件: 1. 趨勢上漲、第一根下跌、第二根上漲、第三根上漲 2. 第三根收盤價 \geq 第一根 $1/5$ 的價格 3. 第二根開盤價 \leq 第一根 $1/2$ 的價格。



BearishHarami 需符合以下要件: 1. 趨勢上漲、第一根上漲、第二根下跌 2. 第二根收盤價 $>$ 第一根開盤價 3. 第二根收盤價 $>$ 第一根開盤價。



BullishHarami 需符合以下要件: 1. 趨勢上漲、第一根上漲、第二根下跌 2. 第二根收盤價 > 第一根收盤價 3. 第二根收盤價 < 第一根開盤價。



在設定交易訊號的條件時，需面臨相似性與數量的抵換關係，條件設定的愈嚴格雖然反撲或反彈的力道愈強，但符合該條件的樣本就愈少，之後在模形訓練時會面臨 label 過少，資料不平衡的問題；相反的，條件設定愈寬鬆，雖然符合該條件的樣本會變多，愈好訓練模形，但模形所預測並推薦的形態不一定會是有力的交易訊號。

最終偵測結果顯示，符合 MorningStar 的區段共 44 個，符合 EveningStar 的區段共 31 個，符合 BearishHarami 的區段共 39 個，符合 BullishHarami 的區段共 30 個。為了讓之後模型的訓練資料能夠平衡，將資料中分成 5 種型態: 1. 符合 MorningStar 的區段 2. 符合 EveningStar 的區段 3. 符合 BearishHarami 的區段 4. 符合 BullishHarami 的區段 5. 不屬於任何形態的區段。每種型有 30 個區段，每個區段有 10 筆樣本資料，總共 1500 筆訓練資料，

2.3 GASF 轉換

使用 Gramian Angular (GAF) 轉換，將一維的時間序列資料轉換為二維的圖像資料。將 (150, 10, 4) 的時間序列資料轉成 (5, 30, 10, 10, 4) 的圖像資料。

Normalizing the data into the interval $[-1, 1]$ or $[0, 1]$

$$X = \{x_1, x_2, x_3, \dots, x_n\} \quad X' = \{x'_1, x'_2, x'_3, \dots, x'_n\}$$

Transforming the rescaled time series into the polar coordinate system

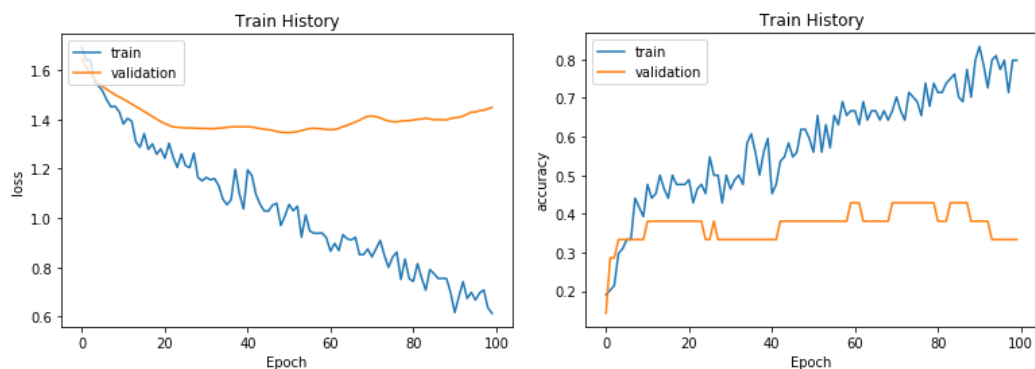
$$\phi = \arccos(\tilde{x}_i), -1 \leq x'_i \leq 1, x'_i \in X'$$

$$\text{GASF} = \begin{bmatrix} \cos(\varphi_1 + \varphi_1) & \cos(\varphi_1 + \varphi_2) & \dots & \cos(\varphi_1 + \varphi_n) \\ \cos(\varphi_2 + \varphi_1) & \cos(\varphi_2 + \varphi_2) & \dots & \cos(\varphi_2 + \varphi_n) \\ \dots & \dots & \dots & \dots \\ \cos(\varphi_n + \varphi_1) & \cos(\varphi_n + \varphi_2) & \dots & \cos(\varphi_n + \varphi_n) \end{bmatrix}$$

2.4 訓練 CNN 模型

將訓練資料集分成(150, 10, 10, 4)的影像資料與(150,)的 Labels，使用 CNN 模型的預設參數。驗證資料的 loss 與 accuracy 分別為 1.0511 與 0.4444。

| Layer (type) | Output Shape | Param # |
|--------------------------------|--------------------|---------|
| conv2d_5 (Conv2D) | (None, 10, 10, 32) | 544 |
| max_pooling2d_5 (MaxPooling2D) | (None, 5, 5, 32) | 0 |
| conv2d_6 (Conv2D) | (None, 5, 5, 32) | 4128 |
| max_pooling2d_6 (MaxPooling2D) | (None, 2, 2, 32) | 0 |
| dropout_5 (Dropout) | (None, 2, 2, 32) | 0 |
| flatten_3 (Flatten) | (None, 128) | 0 |
| dense_5 (Dense) | (None, 128) | 16512 |
| dropout_6 (Dropout) | (None, 128) | 0 |
| dense_6 (Dense) | (None, 5) | 645 |
| Total params: 21,829 | | |
| Trainable params: 21,829 | | |
| Non-trainable params: 0 | | |



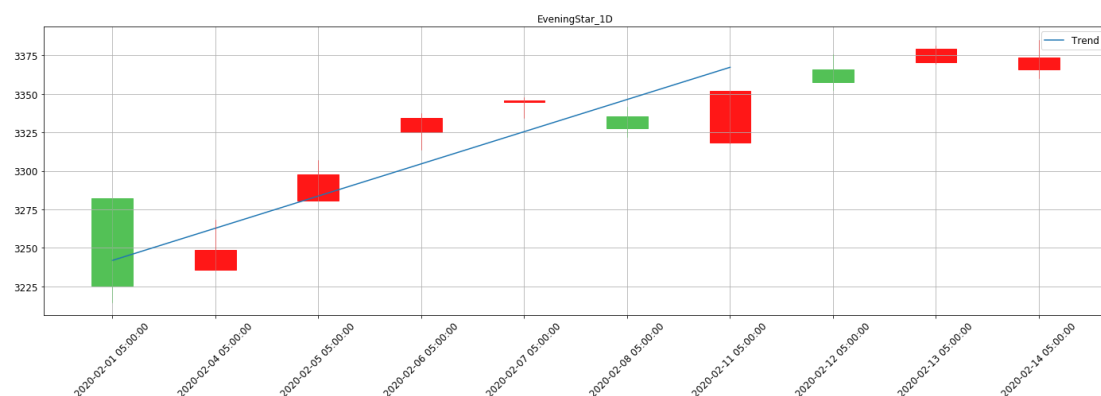
2.5 載入即時資料 & 預測即時資料的形態

即時資料的時間尺度為「分」，載入一筆即時資料與前 9 筆歷史資料，共 10 跟 K 棒進行型態預測。將美國時區時間轉台灣時區時間 UTC+8，夏令時將台灣時區

開盤時間往前調整一小時。

由於偵測的時間尺度為「日」，所以在交易時段內，除了收盤當下會使用當日收盤的 ohlc 加上前 9 筆歷史資料進行偵測，其餘時間都使用前一日收盤的 ohlc 加上往後數前 9 筆歷史資料來進行偵測；非交易時段則不做任何偵測。

如圖所示，在 2020-02-14 時，模型辨識該筆即時資料的形態為 EveningStar。



三、 檢討

從模型的正確率偏低以及預測結果來看，主要原因在於訓練資料量過少的問題，時間範圍從 2015/2/18 到 2020/2/13，總共 1258 筆，經過形態偵查和 GASF 轉換後僅剩 150 筆，對於訓練 CNN 模型而言過於稀少。而這本身的問題在於時間尺度太大的關係，因為礙於無法找到 S&P 500 的分資料，僅能以日資料代替，所以才造就資料量過少的問題。如果能找到標的資產更小的時間尺度資料，相信不論對模型準確度或是投資人交易策略方面都能有顯著的改善與幫助。另外，在串流即時資料方面還需解決夏令與冬令交會的時段以及遇到特殊假期休市的情形。