

# 類神經網路

## Project 1

姓名：洪偉倫  
系級：資工 111  
學號：40747026S

### 1. 電腦硬體

- CPU: Intel i7-9700
- RAM: 32G
- GPU: RTX 3060 O12G

### 2. 電腦軟體

- OS: Ubuntu 18.04.5
- NVIDIA Driver: 460.67
- CUDA: 11.0
- cuDNN: 8.0
- torch: 1.7.1 + cu110
- torchvision: 0.8.2 + cu110
- torchaudio: 0.7.2
- scikit\_learn: 無指定版本
- matplotlib: 無指定版本
- numpy: 無指定版本
- pandas: 無指定版本
- tqdm: 無指定版本

### 3. 資料預處理

由於我認為以數值進行訓練可能會成效不彰，測試的結論也確實是如此，於是我從原始資料做了一些運算，增加了幾欄我認為對股票預測有所幫助的數值。以下為列表：

- Length: High - Low, 屬於數值。
- BarLength: Close - Open, 屬於數值。
- RodLength: High - max(Open, Close), 屬於數值。
- PinLength: Low - min(Open, Close), 屬於數值。
- OpenRatio: 今日 Open  $\div$  昨日 Close, 屬於比率。
- HighRatio: High  $\div$  Open, 屬於比率。
- LowRatio: Low  $\div$  Open, 屬於比率。
- CloseRatio: Close  $\div$  Open, 屬於比率。
- RiseFall: (Close - Open)  $\div$  Open, 屬於比率。
- BarRatio: BarLength  $\div$  Length if Length  $\neq 0$  else 0, 屬於比率。
- RodRatio: RodLength  $\div$  Length if Length  $\neq 0$  else 0, 屬於比率。
- PinRatio: PinLength  $\div$  Length if Length  $\neq 0$  else 0, 屬於比率。
- RSI\_5: 5 日內相對強弱指標。
- RSI\_10: 10 日內相對強弱指標。

加上此幾欄資料，原本的資料將從 5 欄 (Date、Open、High、Low、Close) 擴展為 19 欄。並且融合台股自身資料與大盤資料，輸入資料將來到 37 欄 (19  $\times$  2 - 重複的 Date)。

#### 4. 模型架構

我挑選的模型架構為 Informer，是一個基於 Transformer 架構改良而來，於 2020 年發表。

Informer 目前在 paper with code 網站上的 Time Series Forecasting 項目中佔據所有 Dataset 的準確率第一名。

此為 Informer 論文在 Paper with code 網站上的網址：

<https://paperswithcode.com/paper/informer-beyond-efficient-transformer-for>

此架構運作模式為 Encoder-Decoder。

此架構訓練時需要的輸入與答案較為特殊，在此以 train\_X、train\_Y 表示輸入與答案，並以 A、B、C 表示某時間長度。

train\_X 是長度為 A+B 的序列資料，而 train\_Y 則長度為 B+C 的序列資料；其中的特別處在於，C 才是我們要預測的解答部份，而 A、B 可以自由指定。

所以假設我今天我要預測接下來 5 天內的資料，我將會餵給模型長度為 A+B 的 train\_X 和長度為 B+5 的 train\_Y。

Encoder 會先經由神經網路將 train\_X 濃縮，並讓 Decoder 學習 train\_Y 和 train\_X 之間的關聯性。如此一來，Decoder 將可以窺探過往的資料，獲得一個更長期的概要資訊，強化預測未走勢的能力，也就是推算 train\_Y 最後面長度為 C 的解答的能力。

此架構預測時，一樣是餵給模型長度為 A+B 的 train\_X 與長度為 B+C 的 train\_Y，但模型會無視 train\_Y 最後長度為 C 的部份的數值資料 (不會無視日期資料)，所以該處資料可以隨便指定數字 (例如 0)。但在填充 C 部份時要注意，因為這個架構是會處理時間並將時間納入訓練參數的，所以 C 部份的日期資料仍要正確。例圖如下：

	A	B	C	D	E	F	G
	Date	SI_Open	SI_High	SI_Low	SI_Close	SI_Length	SI_BarLength
288	2021/3/15	16256.58	16281.91	16194.91	16249.33	87	-7.2
289	2021/3/16	16253.93	16340.66	16244.99	16313.16	95.67000000000001	59.22999999999999
290	2021/3/17	16312.98	16349.21	16166.35	16215.82	182.85999999999999	-97.15999999999999
291	2021/3/18	16264.61	16410.02	16264.61	16287.84	145.41	23.22999999999999
292	2021/3/19	16186.46	16186.46	16022.17	16070.24	164.28999999999999	-116.21999999999999
293	2021/3/22	16065.5	16235.63	15983.77	16189.22	251.85999999999999	123.71999999999999
294	2021/3/23	16250.2	16351.38	16166.29	16177.59	185.08999999999998	-72.61000000000000
295	2021/3/24	15994.96	16125.54	15967.95	16032.12	157.59	37.16000000000000
296	2021/3/25	16010.54	16146.58	15944.96	16060.14	201.62000000000001	49.59999999999999
297	2021/3/26	16140.37	16325.62	16140.37	16305.88	185.25	165.50999999999999
298	2021/3/29	0	0	0	0	0	0
299	2021/3/30	0	0	0	0	0	0
300	2021/3/31	0	0	0	0	0	0
301	2021/4/1	0	0	0	0	0	0
302	2021/4/2	0	0	0	0	0	0
303							

在本次專案中，A 值我設為 20，B 值我設為 40，C 值則是 5。

這個架構確實十分強大，但畢竟不是專門預測股票，而是預測週期性長時間序列的模型。因此，原始的移動平均法或是簡單粗暴的幾層神經網路能勝過他也不奇怪。我只是想藉機學習一下比較複雜一點的新技術。

## 5. 一些訓練圖表

