深度強化學習與應用

期末報告 group 3

資科四 106703055 黃浩瑋

資管碩一 109356011 簡琬玲

資管碩一 109356024 蘇品維

目錄

[一、.資料前處理: 2](#_Toc75383131)

[(一)使用資料集: 2](#_Toc75383132)

[(二)資料預處理: 2](#_Toc75383133)

[二、期中進度反思與分工 2](#_Toc75383134)

[(一)期中分工: 2](#_Toc75383135)

[(二)Deep Q learning部分: 3](#_Toc75383136)

[(三)Stable base line部分: 4](#_Toc75383137)

[(四)最後決策: 4](#_Toc75383138)

[(五)最終分工: 4](#_Toc75383139)

[三、針對action1的調整--手動挑選基金: 5](#_Toc75383140)

[(一)確立挑選方式: 5](#_Toc75383141)

[(二)挑選指標介紹: 5](#_Toc75383142)

[(三)挑選步驟: 9](#_Toc75383143)

[(四)加入消息面 : 11](#_Toc75383144)

[(五)加入情緒分數: 12](#_Toc75383145)

[四、針對action2的調整--十檔基金的比例分配訓練 13](#_Toc75383146)

[(一)環境設置 13](#_Toc75383147)

[(二)模型訓練 16](#_Toc75383148)

[(三)Final Result (2019/01~2020/06) 19](#_Toc75383149)

[六、未來展望: 20](#_Toc75383150)

[七、參考資料 20](#_Toc75383151)

# 一、.資料前處理:

## (一)使用資料集:

本堂課助教所給的2015~2019的8000多檔基金淨值資料，以及相關基金屬性資料。

## (二)資料預處理:

1.首先將以下2015~2019NAV檔案，以垂直合併的方式將資料嫁接在一起，並依照日期排列。即可得出每日基金淨值的資料，若該日無基金淨值資料，則視為基金價格無變動，因此價格會與上一個日期時間點所呈現的基金價格相同。

2.將其他基金屬性資料以對應的Index欄位加以利用left join的方式以基金淨值資料，加以水平合併。而發放配息的欄位若為NAN，則以0填入。視為該日不發放股息。

3.利用pandas .piviot\_table的方式將DataFrame的column轉變成基金名稱(aggregate function=’first’)，而row變成日期，裡面的內容則是該基金對應到該日期的屬性值。我們採用的屬性有發行國家，風險等級、基金淨值、基金變動率、以及是否配息這五種屬性，因此會產生五張DataFrame。

4.利用np.stack將五張表格交疊起來，形成三維資料(shape=(天數、基金檔數、屬性種類個數))，並reshape成二維的txt檔輸出。之後模型若需要使用state的話，只需匯入txt檔，還原原本的維度，並利用numpy array輸入三維index的方式抓取資料即可。

# 二、期中進度反思與分工

## (一)期中分工:

主要是在釐清問題，所以利用DQN以及Stable base line進行簡單版本的實作，並且看看哪個效果較佳。

|  |  |  |
| --- | --- | --- |
|  | 研究 | 實作 |
| DQN | 簡琬玲 | 蘇品維 |
| Stable base line | 黃浩瑋 | |

## (二)Deep Q learning部分:

1.方法:

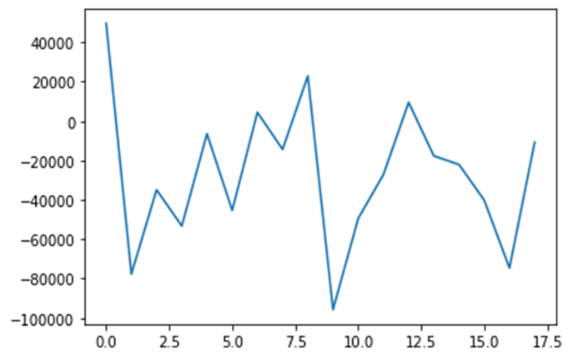
期中進度報告時，我們利用Deep Q learning Network 將預處理完成資料的2292檔基金，並且利用每日所變動sharp ratio 作為reward function 之設計，state 部分則是取發行國家，風險等級、基金淨值、基金變動率、以及是否配息進行模型的訓練。模型使用epsilon greedy =0.3。我們每天會固定挑選十檔基金，並且設立兩種action:

Action1:選擇2292檔基金中的哪十檔作為資產配置的基金標的(離散型action)

Action2:從挑出來的十檔基金中，每一檔基金所配置的權重，並且根據我們目前剩餘的預算，計算出在這樣的預算限制以及權重分配之下，每天能購買的基金張數，以及比例分別為何?基金的權重以每日的方式進行調整。

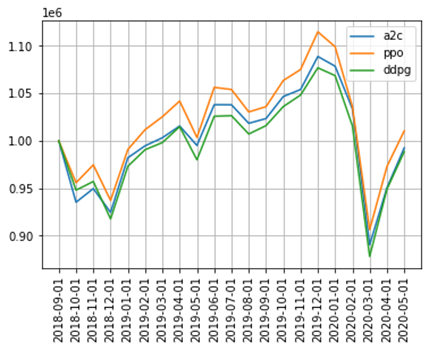
2.初步結果:

下圖為模型的training data的每日獲利分布(x軸為進行天數，y軸為獲利)，可以看出獲利極為不穩定，min max draw down 高達123000元，且獲利大概只有9437元。我們推測從2292檔基金中挑選10檔基金這個問題，在配合我們所提供的state資料之下，模型的任務過於複雜。再加上權重比例的的分配屬於連續型的action，而DQN的action屬於離散型，可能會導致模型action上的調整出現問題。



## (三)Stable base line部分:

以隨機方式挑選十檔基金，state的部分則是只挑選基金淨值帶入三種模型(A2C,PPO,DDPG)進行訓練，訓練的結果如下圖可見 ，最終獲利與DQN差不多，但min max draw down 比DQN來的穩定，但還是有過大的情形。



## (四)最後決策:

1.針對上述action1調整:

考量到action1的部分對於模型來說過於複雜，因此我們改成利用金融市場其他額外衡量風險以及報酬的基本面資料，進行十檔基金的手動挑選。以及加入總體經濟相關的消息面文字資料的word vector 資料、金融相關技術指標當成模型state的部分進行訓練。

2.針對上述action2的調整:

將上述DQN之訓練權重的方法改成Stable base line，並且沿用DQN的output方法，利用softmax 將結果限制在 Σ output=1，並且加入action1調整所額外萃取的state(基本面、消息面、基金淨值)資料，增加訓練state的維度。演算法的部分則是使用ddpg，較為適合訓練連續型output的演算法進行模型的訓練。

## (五)最終分工:

|  |  |  |
| --- | --- | --- |
|  | 負責內容 | 負責人 |
| action1調整 | 金融指標研究計算、消息面的爬取與特徵轉換 | 簡琬玲、蘇品維 |
| action2調整 | 模型建置(十檔基金購買比例的權重分配)。 | 黃浩瑋 |

# 三、針對action1的調整--手動挑選基金:

## (一)確立挑選方式:

為了能夠更確立我們未來的發展方向，我們參考了胡毓忠老師的學生所發表的畢業論文。裡面的未來展望提及，可以利用相關金融指標，例如:KD、RSI指標加入模型的State進行模型的優化，因此我們決定對於相關金融基本面指標做出一些相關的研究。

經評估及研究後，我們決定加入KD、RSI兩個與「報酬」相關的技術指標，以及價格變動標準差、最大回撤 (Maximum Drawdown)、最大下限(Downside Deviation) 、Beta係數這四個與「風險」相關的技術指標作為挑選基金的衡量方式。而我們參考較多的風險指標是因為訓練資料集大多屬於多頭的情況。在這情形下，若考慮太多獲利報酬指標，未來若遇到下跌的情形，整體資產配置的min max draw down可能會遇到過大的問題。以下分別介紹各項技術金融指標的意義。

## (二)挑選指標介紹:

1.KD

KD指標的全名為隨機指標(Stochastic Oscillator)，是由K值和D值所組成的兩條線圖，適合作為短線研判買賣點之用。

當KD數值越高代表個股的收盤價越接近最近N天的最高價，反之當KD數值越低則代表個股的收盤價越接近最近N天的最低價。

K值和D值都代表股價的變化速度，其中K值又稱快線，能較快速地反應股價變化；D值又稱慢線，相較K值而言對股價變化的反應較不靈敏，因此，一般認為當K值>D值時，股價為上漲行情，適合做多，當 D值>K值時，則屬於下跌行情，適合做空，兩線的交叉點也被當作買賣時機的判斷方式之一。

而在計算KD值之前，必須先求得RSV。RSV的中文全名為未成熟隨機值(Row Stochastic Value)，它代表著當天股價在最近N天內是相對強勢或弱勢，通常習慣將N設為9。

我們先算出每檔基金的RSV再求得K值和D值，接著統計每檔基金K值>D值的總天數，作為後續挑選基金的評估方式之一。

RSV、K值、D值的計算方式:

RSV = (今日收盤價 - 最近九天最低價)/(最近九天最高價 - 最近九天最低價) \* 100

今日K值 = 2/3 \* (昨日K值) + 1/3 \* (今日RSV)

今日D值 = 2/3 \* (今日D值) + 1/3 \* (今日RSV)

2.RSI

RSI指標(Relative Strength Index , RSI)又稱為相對強弱指標，主要用來衡量在過去一段時間(N日)內，股價買盤與賣盤的相對強弱比例，其中考慮了股價變動的四個因素: 上漲/下跌天數和上漲/下跌幅度，可用來評估股價超買或超賣的情況。

當RSI越高時，代表買方力道越強，預期股價上漲相對多，一般認為大於80時表示市場已過熱，須注意反轉(下跌)現象；而當RSI低於20時，則代表市場過冷，股價很可能準備逆勢上漲；通常會以50作為區隔。

至於「過去一段時間」的天數，一般較常見的為6、12、14日，我們將N設為12，計算每檔基金的RSI大於50且小於80的總天數，即買方力道較強勁、平均上漲幅度大於平均跌幅的總天數，作為後續挑選基金的評估方式之一。

RSI計算方式:

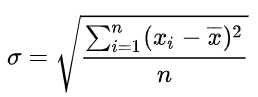
RSI = 前N日漲幅平均值/(前N日漲幅平均值 + 前N日跌幅平均值) \* 100

3. 價格變動標準差

標準差代表著基金淨值在一段時間內的波動情況，一般而言，當標準差愈大，表示淨值的漲跌較劇烈，風險程度也較高。

我們先計算每檔基金的平均淨值，進而求出價格變動標準差，作為衡量每檔基金波動風險的方式之一。

數學式:

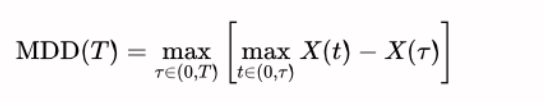


4. 最大回撤 (Maximum Drawdown)

這衡量了投資組合從高峰到低谷的最大跌幅。此指標需要考慮較長期的投資時間，其中應包含空頭市場。如果僅是查看資產上揚期間的最大跌幅，此數值會只是個很小的百分比，這對於分析空頭時期的投資組合風險是沒有用處的。而通常股票型基金的最大回撤會大於債券型基金，最大回撤反映了買入某基金後基金可能出現最糟糕的情況，當我們在挑選基金時，若最大回撤指標超出了我們的承受範圍，那就必須謹慎考慮是否要將它納入我們的資產配置中了。

* 投資組合連續4年產生穩定報酬5%，其最大回撤為0.0%。
* 投資組合的報酬是 +5%，接續 +10%、 -10%，最後 -5%，其最大回撤為14.5%
* 投資組合的報酬是+4%，接續 +5%、 +6%，最後 +5%，其最大回撤為0.0%，因為在此期間都沒有負報酬。

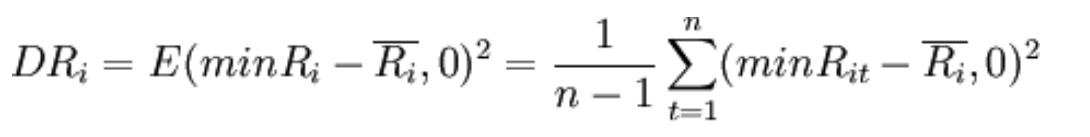
數學式:



5. 下限風險 (Downside Deviation)

將標準差視為組合的總風險，描述組合收益率高於或低於平均收益率的波動幅度。但在實際的投資中，投資者往往會對收益和風險有一個權衡，與收益率高於基準部分所帶來的收益回報相比之下，投資者可能更關心低於基準部分的收益下跌風險，即所謂下限風險(Downside Risk)。當投資組合的收益率服從常態分配時，只有用下限風險來衡量投資組合風險，因為此計算也考量加入了時間序列的考量，才能精準的考慮到每支基金隨著時間的標準差變化，才能幫助投資和出正確的決策。

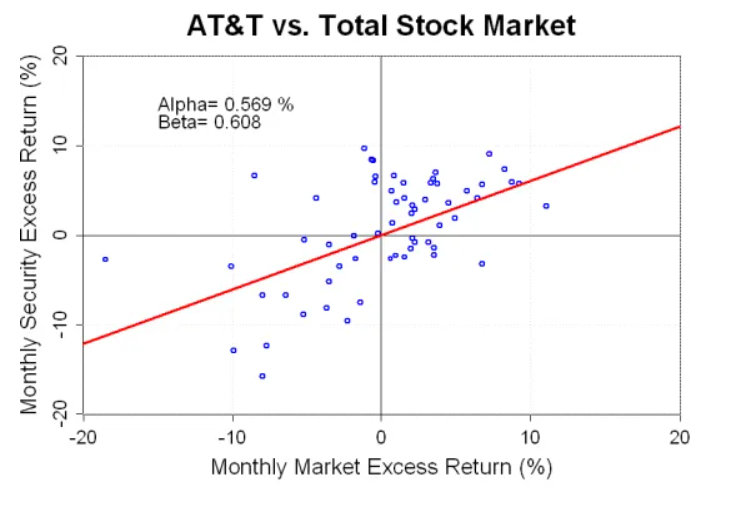
數學式:

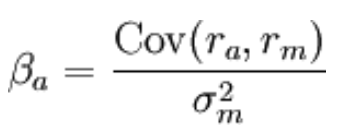


6. Beta係數

是一種評估證券系統性風險的工具，用以度量一種證券或一個投資證券組合相對總體市場的波動性，雖然系統性風險無法透過基金投資組合加以避免，但是此指標可以間接的衡量每一支基金相對於大盤的波動率，因為考量大盤大波動相對於相對於單支基金的平均波動率較小，再加上以往的基金市場大多呈現多頭市場，若用此方式估算，多少會形成一個風險與獲利綜合考量的指標。而在這次的資料集中，大盤資料指的是所有基金淨值的價格平均。

數學式: 將整體市場報酬率、對單一資產報酬率做回歸分析時，Beta 值就是斜率的係數





其中

ra=該基金的收益

rm=整體市場(平均收益)

σm^2=整體市場收益的變異數。

## (三)挑選步驟:

(1)先將上述所提及的金融面指標利用python加入衍伸欄位中。

(2)將2000檔基金從該指標的風險或報酬中由好到壞進行排序，並且抓取前300筆作為篩選標的。

(3)將風險指標與報酬指標分別作出交集，得出經由多種風險指標與多種獲利指標的綜合排名最高的基金

(4)將分為兩種方式進行挑選:

1.風險為主、獲利為輔的方式:將由風險指標衡量過後的指標也到獲利指標進行衡量，若獲利能力在500名之外者進行剔除。

2.獲利為主、風險為輔:由獲利指標衡量出來的基金也以同樣的方式衡量風險。

(5)最後得出最終十檔基金標的物，

(6)挑選結果呈現:

最終我們挑選出來的基金標的為基金代碼'762', '1908', '1762', '1912', '2110', '929', '4407', '4405', '3148', '4163'這十檔基金標的。為了在進行整體風險與報酬之平衡，我們將上述所挑選出來的基金淨值變動進行視覺化，從曲線中的波動進行風險與報酬的確認，下圖為視覺化之結果，可以明顯的看出3148,929,2210這三檔股票是採報酬為主，風險為輔的衡量模式所挑選出來的基金，他的波動會相較於其他三檔來的大。而其他則是採用以風險為主，報酬為輔的方式所挑選出來的標的，幾乎沒有甚麼波動。整體的投資組合報酬的曲線為下圖的portfolio曲線，可以看出獲利指標挑選出來的基金在損失時，對於我們的投資組合來說的損失是較少的。



## (四)加入消息面 :

因為市場面的系統性風險是較無法透過投資組合的分配來進行避險的。所以我們決定利用消息面將總體經濟的面向納入我們的模型，希望模型能夠學習到以上十檔基金，哪一檔基金可能會較為受到總體經濟層面所影響，並希望能藉由權重分配的學習，再更降低整體基金報酬的風險，使得我們的獲利能夠更趨於穩定。

1.挑選資料來源介紹:

所採用的消息面資料為「第一銀行」官網所提供的總體經濟面新聞標題。我們先透過chrome driver 將非假日的日期，經由網頁控件一一逐步篩選，並且爬取從 2015-01-23~2019-12-31 的新聞標題，取出每天熱度前十名的新聞標題進行詞向量的轉換，以下將說明詞向量的轉換方式。

2.消息面資料處理:

由於模型只能處理數值型資料，因此首先必須將新聞標題進行詞向量轉換。我們利用結巴(jieba)套件中的精準模式將每則中文標題進行斷詞，即把一句話拆成許多詞彙，並且去除停用字，例如: 並且、不但、假如、以免...等等，只留下每則新聞較具有意義的詞。為了更準確地取出重要詞彙，我們還使用由Stanford NLP 團隊針對python所開發出nlp分析套件-Stanza，將所有屬於名詞、動詞以及形容詞的詞彙挑選出來，作為最終每則新聞標題的代表。

接著，我們透過word2vec模型工具，把所有文字轉換成向量型態表示，而由於輸入的詞彙總數量多，每則詞彙會變成由100個數值表示。

範例:

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 新聞標題 | 斷詞結果 | Stanza詞性挑選 |
| 2015/3/12 | 富邦人壽去年獲利再奪冠；今年徵才6800人 | 富邦 人壽 去年 獲利 奪冠 今年 徵才 人 | 人壽 去年 獲利 奪冠 今年 徵才 人 |

|  |  |
| --- | --- |
| 詞彙 | 詞向量轉換 |
| 去年 | [-3.58251512e-01, 2.71741748e-01, -1.75890811e-02, -7.39504620e-02, … , -1.02303743e-01, 1.95837572e-01, -9.87455249e-02] |

## (五)加入情緒分數:

根據政大資管系姜國輝教授的論文--股市趨勢預測之研究 -財經評論文本情感分析所述:「消息面的情緒會間接的影響到財經市場的走勢」，因此我們決定將新聞標題的文字轉換成情緒分數，也丟到模型內訓練。我們利用 google cloud api 所提供的nlp套件。此套件將會把新聞標題的中文文字轉為由-1~+1的分數，分數越高代表此文字所透露出來的正向情緒越高。

範例:

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 新聞標題 | 熱度 | 情緒分數 |
| 2015/2/4 | |  | | --- | | 復航班機墜落基隆河 已知13死28人送醫 | | 590 | -0.4 |
| 2015/1/23 | |  | | --- | | 去年商業營業額創歷年新高，年增2.7% | | 2428 | +0.9 |





我們再將一天所獲得的十則情緒分數進行加總，並算出每天新聞的情緒加總作為當日的情緒分數。

最終我們將上述所提到的資料面相加以整理，並且納入模型的state進行training。

# 四、針對action2的調整--十檔基金的比例分配訓練

## (一)環境設置

1.Action

這裡的Action指的是我們對資產權重所做的動作，有可能是離散的針對各個基金做出買、賣或是持有的三項動作，又或者是建立在連續的空間下，這裡我們所選擇的是使用連續空間下的動作分布。

除了需要先決定我們要採取的動作是否連續外，其所代表的意義也是重要的，我們原先考慮的是將其作為我們基金權重的改變幅度，因此它的範圍勢必是[-1,1]，因為其不可能超越100%的幅度，然後這種做法存在著些許問題。

首先，是控制幅度的問題，不能超越100%的限制十分清晰，因此這倒不是問題，然而卻也不存在小於0%的基金權重，因此需要特別注意不得扣超過原先所持有的比例。

其二，為第一個問題的延續，因我們需要於上千基金中選出我們要投資的十檔基金，在這部分，除了要注意負值的權重外，基金權重總和不得超過100%也是一項大問題，我們是否會難以回收我們投資在一檔基金上的資金，導致其他的投資行為受限也是一大隱憂之一。

綜合上述，我們選擇將Action表示為我們每次交易的新基金權重，並且可以輕易用Softmax將其限制在[0,1]之間，並且全部加總為100%，最重要的事也不會存在難以回收的問題出現，只需將其權重化為0即可。

2.State & Observation

State與Observation的定義也是十分重要的，不過這裡我們單純地將State作為Fully Observable，先使這兩個因素化為同一因素。

而關於State的定義，首先，我們採用了基金淨值前一日的變化率，不過為考慮到穩定性，降低被短期變動影響的可能性，我們額外加上了昨天之於三天前以及五天前的變化率。

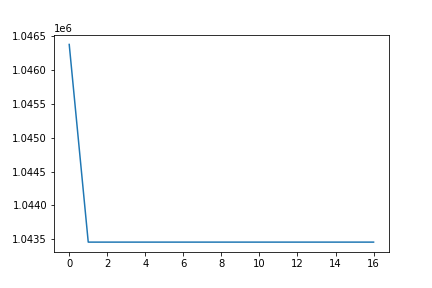
另外，因為為避免每日交易過度頻繁，我們決定每五天(相當於一周扣除假日)進行一次資產重配置，以降低Transaction cost所帶來的衝擊，不過若我們單純每五日才進行依次資產重配置，而可觀察因素也只有昨日之於昨天、前三天以及前五天的變化率，這些少量因素也許會遺漏掉一周內基金的重要改變過程，因此，我們除了昨天以外，也額外補充了前五天的所有基金變動率。

最終，若只看基金淨值變動率將會讓此任務變成單純的數值預測，為此，我們也加入了許多重要的金融指標如：DD、MDD、RSI等，運用金融的角度協助我們進行資產重配置的判斷，另外，需特別注意專案所要求的75%股債比例限制。

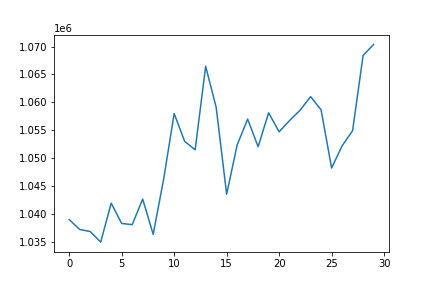
3.Reward Function

Reward function的制定決定了我們告訴機器他的終極目標為何，制定Reward function需非常小心，因為往往我們認為的只限於我們所看到的或是額外增加了自己認為的常識，然而當機器搞錯我們的意思時，我們將難以達到所希望獲得的效果，在實驗過程中，我們也曾遭遇過這一類的問題，因此我們也嘗試了許多的Reward Function：

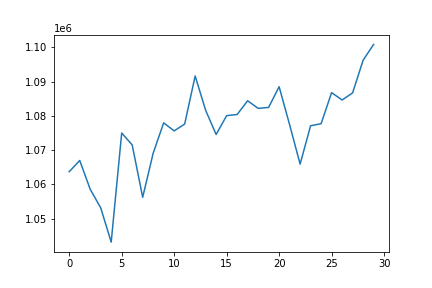
* Total Asset Amount

首先，我們需先確認機器該學習的終極目標為何者？最直觀的想法即為我們資產的總額，極大化我們的總資產應該就是我們的主要目標，然而，我們卻得到了如右圖的結果，此為資產總額經過不斷訓練的過程中在驗證資料集的變化。

* Logarithm of the Marginal Portfolio Profit

我們認為造成上述現象的原因是因為整個Reward function 的值過大，相比之下，每次的回報變動十分地小，導致機器無法進行學習。為此，我們更改了我們的Reward function，將其改為新資產之於舊資產的改變比率，特別再取Log函數，以讓回報的值聚集避免過於稀疏。

* Marginal Portfolio Profit Changed Ratio

而上述兩個函數皆看重總資產的成長，然而我們認為若極大化每次的回報也會相當於極大化總資產，因此我們又嘗試了另一種作法，即為利潤之於舊資產的改變百分比率，而這次的實驗結果為全部最好的。  


不過上述三個Reward function皆只考慮利潤，而並未考慮風險，這會讓我們的結果具有潛在的危險，因此我們最終選擇了下列兩個Reward function：

* Marginal Portfolio Profit Changed Ratio + MDD

不過，我可以明顯觀察到若只單純考慮最大化每次交易日所獲利潤，我們勢必會遺失風險控管的重點，因此我們在上述三個Reward functions中挑選出表現最佳的，並且額外加上MDD的因素，以將風險也納入考量之中。

* Sharpe Ratio

除了上述自製的回報函數外，我們也嘗試了同時考慮了利潤以及風險的指標— 夏普率，夏普率考慮了每次交易所獲利潤的平均以及其標準差，我們所求的是低風險高穩定的高回報，而這似乎與我們自製的回報函數在某種程度上不謀而合。

4.Transaction cost

* Clip action changes

為了更貼近真實情況以及符合專題需求，我們需要特別計算交易手續費，交易手續費雖然只有1%或2%，然而若不謹慎處理將會造成很龐大的負擔，為此，我們需特別對此進行處理，我們所做的第一個動作便是把權重改變的幅度給予限制，我們假設在一次交易中合理的幅度上限為5%，否則須承擔過高風險，誘因我們需要鼓勵權重改變以找到最佳值，因此下限定為0.5%。

* Trading Interval

另外，我們可以發現到交易手續費上升的最大因素為「交易次數」，因此我們透過將每日交易改為每周交易，大幅度地降低了交易手續費，不過這種作法算是種Trade-off，畢竟我們拋棄了資產重配置的靈活性，有可能會進一步降低整體總收益，而根據實驗我們也發現若不論手續費的高低，每日交易的總資產也會明顯大於每周交易的，因此，此種作法仍有待改善。

## (二)模型訓練

1. NULL Model

為了比較我們的模型好壞，最基本的效能便是要比瞎猜要好，否則便不需進一步地去理解了，這邊我們先定義我們的Null Model為投資各10%，接下來便不進行任何的資產重配置了，根據實驗，初始資產為1,000,000的狀況下，我們得到的答案為974,818，除此之外，當然還需特別扣除交易手續費的部分，不過，這個還算不錯的結果或許可以歸功於十檔基金的前置挑選。

1. DDPG

首先，針對這次的專題設計，我們特別專注在DDPG之上，專注在如何克服DDPG所帶的問題以及不穩定性，而並未刻意與其他模型進行比較，除此之外，我們的實驗是透過OpenAI Gym以及StableBaseline3進行架設。

不過，為甚麼我們要選擇使用DDPG作為我們的演算法呢？這乃是因為一開始我們的計畫是想使用DQN進行訓練，然而DQN明顯的缺點即為無法應用在連續動作空間，而DDPG又相當於可以應用在連續動作空間的DQN。另外，最為重要的是我們的專案不只是需要連續動作空間，其動作空間的複雜度還挺高，而DDPG的特色—Deterministic Policy Gradient又具有須採樣的數據樣本相對少的優勢，即可增快我們訓練效率。

* Setting (default by StableBaseline 3)

1. Policy: mlpPolicy
2. Loss: mean-squared Bellman error (MSBE)
3. Optimizer: adam optimizer

* Exploration Problem (Noise)

DDPG是使用策略梯度的方式進行更新的，然而最重要便是他策略的性質—「確定性策略」，這表示當我們給予他相同的State時，我們將會得到一樣的動作，這樣的結果是極大化Exploitation，然而Exploration卻消失殆盡，極容易陷入局部極大值(極小值)，難以求得我們真正想要的答案，因此，我們需對此做出一些調整。

針對降低探索的問題，最直觀的作法便是讓他無法按照原計畫，被帶領去另一狀況，好以更接近我們想尋找的全體最大值(最小值)，所以我們加入噪音，以干擾Action的變化，這裡對於Noise的設計我們操考了文獻所建議的Ornstein Uhlenbeck Action Noise。

* Replay Buffer

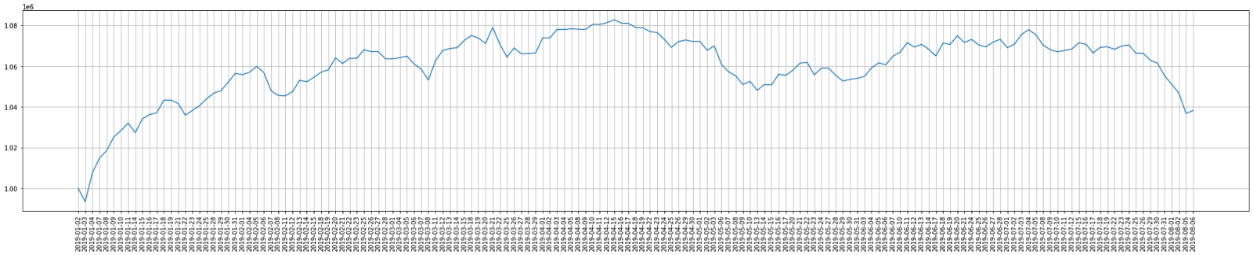
當我們應用DDPG時我們需特別注意一點，因為他是建立在Actor Critic與DQN的結合上，這種如DQN需要透過DNN訓練以得到Q值的作法，不能直接使用RL的數據進行訓練，會礙於其數據間的時間關聯性導致訓練不穩定，為了打斷樣本間的時間關係，DDPG使用了Replay Buffer的技術，以隨機抽樣的方式在Buffer進行學習，然而Buffer Size的制定相當重要，不足的話便會顯得毫無作用，過大的空間也無法有效穩定訓練，根據實驗，我們選擇了Buffer Size=100的作法。

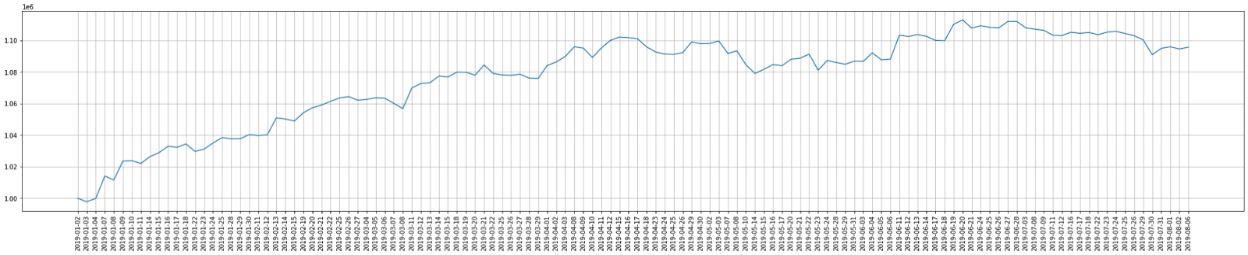
* Tau (Polyak update)

同樣，為了解決上述的時間序列問題，DDPG如DQN也使用了雙重網路的作法(Target Net/ Eval Net)，不過在DQN中，將參數從Eval Net傳至Target Net的做法為「一段時間更新一次」，然而DDPG增加了一個小技巧，多了一個新的參數(tau)，tau的目的即為「每一步都更新一小部分」，根據文獻這種做法能增加訓練穩定性，而根據實驗，我們所選擇的是tau=0.001(應介於0~1)的做法。

五、結果呈現:

1. Original v.s. Improved (Both without transaction cost)





我們可以比較基本模型的成果(上圖)與增加上述改進後的結果(下圖)，兩者皆為(2019/1/2~2019/8/6)的測試資料，我們可以看到當我們增加了許多上述改進後獲利開始相對穩定，並且利潤(不包含手續費)成長了兩倍(8萬 v.s. 4萬)。

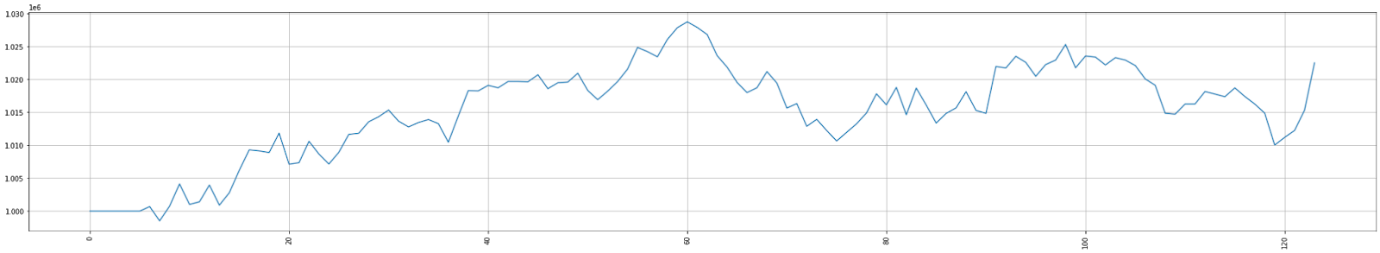
## (三)Final Result (2019/01~2020/06)

1.Our Reward function

(1)CAGR: *0.04553056928141963*

(2)MDD: *-0.018187664454270003*

(3)Process:

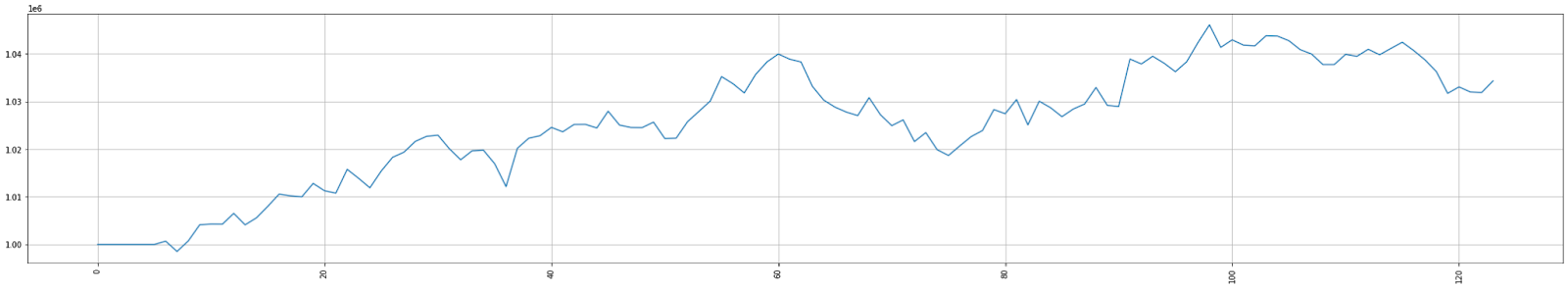


2.Sharpe Ratio

(1)CAGR: 0.069815793761

(2)MDD: *-0.020469415070560548*

(3)Process:



# 六、未來展望:

1 我們這次專案的作法為手動挑選10檔基金，利用ddpg進行10檔基金在每次下單前的權重分配為何，未來若能利用離散型action的模型(ex:DQN)進行基金的挑選，將上述所提到的action1也交給模型進行學習，在實務上或許可以節省許多人力挑基金的決策時間。

2.此次所用的消息面資料為每日的總體經濟層面之資料，未來可以針對在消息面的資料蒐集上加以琢磨。可以將action1所挑選出來的基金進行拆解，判斷出那些基金所包含的公司有哪些，在抓出所有包含那些公司的新聞資料進行訓練。這樣不僅可以參考到整體的系統性風險，也可以針對於各基金文字上的評價做出訓練。

3.DDPG 雖然是基於A2C的DRL框架，可以對於ACTOR所產生的ACTION進行雙重確認，但許多文獻上指出DDPG較不適合用於隨機的場景，且因為policy 的隨機性，雖然我們有利用replay buffer 提升了整體的抽樣，降低了variance，但還是有可能造成在同樣optimal policy 回傳回來的回報值的variance會很大，導致模型訓練的不穩定，未來可以試試看使用Soft-actor-critic的演算法進行訓練。

4.在基金的action的交易頻率上，我們採取的是固定的交易頻率(每五天交易一次)，但其實在實務上，甚麼時候下交易也是對於整體獲利會造成很大的影響，或許後續的專案可以針對交易頻率進行訓練。

# 七、參考資料

1.https://www.itread01.com/content/1548044844.html--

--深度強化學習——連續動作控制DDPG、NAF

2. https://wiki.mbalib.com/zh-tw/%E8%B4%9D%E5%A1%94%E7%B3%BB%E6%95%B0

-- 貝塔係數（Beta Coefficient）

3.https://ah.nccu.edu.tw/item?item\_id=113696

-- 股市趨勢預測之研究 -財經評論文本情感分析Predict the trend in the stock by Sentiment analyzing financial posts

4. https://wealthnavigatortw.com/series/how-to-compare-asset-allocations-using-risk-and-return-measures/

-- 資產配置策略評估：風險及報酬指標

5.https://wiki.mbalib.com/zh-tw/%E4%B8%8B%E9%99%90%E9%A3%8E%E9%99%A9

--下限風險

6. https://rich01.com/fund-indicator-3/

-- 評估基金風險的3大 風險係數 指標：夏普值、標準差、Beta值

7. https://rich01.com/what-is-kd-indicator/

--KD指標原理

8. https://rich01.com/rsi-index-review/

--RSI相對強弱指標原理

9. [https://stable-baselines3.readthedocs.io/en/master/](https://stable-baselines3.readthedocs.io/en/master/--)

--stable baseline 程式參考1

10. https://github.com/AI4Finance-LLC/FinRL

----stable baseline 程式參考2