

應用巨量資料分析與改良粒子群演算法於股票交易策略

Big Data Analytics for stock trading strategy with Improved Particle Swarm Intelligence

陳婷妤¹、劉文卿¹

¹ 國立政治大學資訊管理系 (Department of Management Information Systems, National Chengchi University)

E-mail : timvickie2003@gmail.com, w_liou@nccu.edu.tw

一、中文摘要

股票交易是一錯綜複雜的動態系統，影響交易決策因素多元，加深預測困難度。然股票交易中最重要的事是在交易風險較低的情況下獲利最多，如何降低風險增加獲利程度，一直是相當有趣且實用的議題，在過去文獻有很多學者進行研究。

巨量的歷史執行資料常常蘊含著大量有價值的潛存資訊和知識，近年來由於巨量資料分析興起，帶來資料說話的全新思維解決問題方式，且過去鮮有學者以巨量資料分析進行股票交易策略研究。因此，本研究運用巨量資料分析的資料觀點、平台技術，結合知識發現方法，提出改良式粒子群最佳化演算法，透過股票交易資料、簡單移動平均線技術指標與其黃金死亡交叉決策準則，從中進行知識挖掘，找出股票交易策略新態樣、準則及知識，大幅提升股票投資報酬率，並優於其他策略績效。

關鍵字：PSO (Particle Swarm Optimization), Big Data Analytics, stock trading strategies, stock market forecast, stock market prediction

二、緣由與目的

股票交易是一門複雜的科學，過去文獻研究從解決問題【1】角度探討，有預測指數趨勢方向及波動變化(index fluctuation、trendy direction)【2】、股價預測正確性研究(stock price accuracy、deviation、correctness)【3】、股價變動預測(price movement)【4】、股票績效(predicting stock performance)【5】、股票投資組合選擇(classification of stock)【6】、股票交易策略(stock trading strategy)【7,8】。然再深入剖析股票交易策略研究文獻，發現多以人工智慧結合技術指標，依尋找好的交易決策準則、識別最佳買賣點、判斷趨勢態樣、選擇股票標的等方向，產生股票交易預測策略。而這些文獻均有重要及有趣的發現，但仍有電腦處理資源的限制[17]。若能提供足夠的資源支持資料量較大的運算，將對於股票交易策略的預測能力將有所增進。

另一方面，巨量資料分析(Big Data Analytics)的興起，帶來一些新的觀點、技術與機會，但如何將領域知識、欲解決問題適切結合於巨量資料分析的資料觀點與技術，將是一件相當有挑戰的客題。因此，本研究試圖運用巨量資料分析的資料觀點、分散式運算技術，結合知識發現方法(Knowledge Discovery from Data)【9】，提出改良式粒子群最佳化演算法(Particle Swarm Optimization)【10】，透過股票交易資料、變動天數簡單移動平均線(SMA)技術指標【11】與其黃金死亡交叉決策準則，建構股票交易預測策略，希望從資料觀點挖掘股票交易策略新的態樣、準則，提供投資人進行股票交易決策建議，同時提升平台運算能力。

三、研究方法

本研究運用 Xindong Wu et al.【12】之巨量資料分析框架與 EMC【13】巨量資料分析解決方案，透過知識發現方法【9】與巨量資料分析方式，提出改良式粒子群最佳化演算法，結合股票技術分析之變動天數移動平均線指標【11】，與其黃金死亡交叉決策準則，建構具巨量資料分析特性之股票交易預測策略模型，由股票資料決定佳買賣點交易時機的訊號及決策準則的態樣(pattern)，以獲得最大的股票投資利潤。此外，本研究改良式粒子群最佳化演算法亦與 Kennedy PSO【10】及 MMPSO【8】進行比較探討。

本研究股票交易預測策略模型建構可分為 7 大步驟，步驟說明如次：

步驟 1：定義應用領域問題流程，找出知識發掘重要議題【9】

以股票交易策略而言，終端使用者為投資人，而投資人的股票買賣交易決策決定了買賣點時機與價格，進而影響著股票交易成本及獲利多寡。因此，在股票交易策略的應用領域中，股票交易決策準則、買賣點訊號及態樣將是重要的知識發掘(knowledge discovery)議題。

步驟 2：由知識發掘重要議題中產生資料集合(subset data point)【9】

不同的投資人有不同的股票交易決策準則分析方式，本研究以投資人常用之簡單移動平均線及

其黃金死亡交叉決策準則，定義為本研究識別買賣點訊號的決策準則。該準則常運用短中期及長期的簡單移動平均線進行買賣點訊號偵測，如公式(1)所示。

而其決策準則相關的資料可細分為每日收盤價、SMA 天數、SMA 計算公式等，則為本研究知識發掘之重要資料子集合。

公式(1)

$$signal_k^{\square} = \begin{cases} buy, & \text{if } (SMA_k^{T_1} > SMA_k^{T_2}) \text{ and } (SMA_k^{T_3} > SMA_k^{T_4}) \\ sell, & \text{if } (SMA_k^{T_5} < SMA_k^{T_6}) \text{ and } (SMA_k^{T_7} < SMA_k^{T_8}) \\ hold, & \text{otherwise} \end{cases} \quad (1)$$

步驟 3：將資料子集合進行 ETL(Extract, Transform, Load)

本步驟是整理並取得資料子集合的資料以進行知識發掘，主要是將資料子集合的相關資料欄位進行處理，處理內容包括去除干擾資料 (removing noise)、處理缺漏資料(handling missing data fields)、彙整資料子集合資訊(collecting required information to model)、計算時間及已知變化的資訊(calculating time information and known changes) 【9】。就本研究而言，是從網頁讀取每支股票的每日收盤價、日期整理存入分散式儲存資料庫中。

步驟 4：從資料子集合的資料屬性中，找出影響流程或目標的可調參數 【13】

從資訊子集合之資料屬性中找出重要且可調的參數，以本研究為例，資料子集合為 SMA 決策準則，其資料屬性透過 SMA 公式(1)可得每日收盤價、SMA 天數兩個參數，然其決策準則是由短中期及長期天數 SMA 比較決定買賣點，因此，本研究引用變動天數移動平均線理論(VLMA, variable length moving averages; Gunasekarage, Abeyratna and David, 2001) 【11】，將 SMA 天數變成可調參數，本研究選擇決策準則如公式(1)所示 【8】， T_1 、 T_2 、 T_3 、 T_4 、 T_5 、 T_6 、 T_7 、 T_8 為本研究之可調參數，代表簡單移動平均線之變動天數參數並影響股票交易買賣點決策。

$$signal_k^{\square} = \begin{cases} sell, & \text{if } (SMA_k^{T_5} < SMA_k^{T_6}) \text{ and } (SMA_k^{T_7} < SMA_k^{T_8}) \end{cases}$$

步驟 5：導入適合問題解決的資料分析方法

針對可調參數求解特性及問題解決目標，選擇適當的資料分析方法。就本研究而言，透過步驟(4)得知， T_1 、 T_2 、 T_3 、 T_4 、 T_5 、 T_6 、 T_7 、 T_8 等可調參數，影響著股票交易決策及買賣點的識別，因此，如何找出可調參數的最佳天數解，為本研究知識發掘的重要目標，因此，本研究可調參數的最佳天數求解過程為一最佳化問題。

有關本研究改良粒子群最佳化演算法，簡稱 EPSON (elite PSO) 如公式(2)-(5)所示，其中公式所引用參數 V_{IT} 為粒子速度(velocity for particle X_{IT})， X_{IT} 為粒子位置(the particle position)， P_{IT} 為該粒子之區域最佳解(local best solution)， G_T 為全域最佳解(global best solution) ω 為慣性權重值係數(inertia weight)。 c_1 為粒子自信係數(self-confidence coefficient)，代表粒子自我經驗之最佳解的影響係數。 c_2 為粒子社會信心係數(social confidence coefficient)，代表全體粒子最佳解的影響係數。 r_{i1} 、 r_{i2} 為亂數產生器，其亂數範圍為[0,1]。 n 為本期， $n+1$ 為下一期。

$$V_{IT}^{n+1} = \omega \cdot V_{IT}^n + c_1 \cdot r_{i1}^n \cdot (P_{IT}^n - X_{IT}^n) + c_2 \cdot r_{i2}^n \cdot (G_T^n - X_{IT}^n) \quad (2)$$

$$X_{IT}^{n+1} = X_{IT}^n + V_{IT}^{n+1} \quad (3)$$

$$V_{IT} = \begin{cases} rand[V_{max}, -V_{max}], & \text{if } V_{IT} > V_{max} \\ rand[V_{max}, -V_{max}], & \text{if } V_{IT} < -V_{max} \\ -V_{max} + 2 \times V_{max} \times rand(), & \text{if } |V_{IT}| > V_s \\ V_{IT}, & \text{otherwise} \end{cases} \quad (4)$$

$$X_{IT} = \begin{cases} rand[X_{max}, -X_{max}], & \text{if } X_{IT} > X_{max} \\ rand[X_{max}, -X_{max}], & \text{if } X_{IT} < -X_{max} \\ X_{IT}, & \text{otherwise} \end{cases} \quad (5)$$

公式(6)之投資報酬率(return of investment，簡稱 ROI)為目標函數(fitness function)，來衡量粒子群解之優劣，進而演化到最後得到一組變動天數移動平均線粒子群最佳解，作為股票交易決策參考。投資報酬率公式(6)如下所示：

$$ROI = \frac{GainFromInvestment - CostOfInvestment}{CostOfInvestment} \quad (6)$$

而本研究演算法過程可分為二階段，一為改良的 PSO 演算法主要程式部份，以原 Kenndey PSO 演算法為基礎，再增加一萃取菁英候選解步驟，其主要核心概念是將公式(1)之 T_1 、 T_2 、 T_3 、 T_4 、 T_5 、 T_6 、 T_7 、 T_8 等表移動平均線變動天數的可調參數，亦是本研究之股票交易策略解予以精煉，以找出更好的解。

公式(1)之 T_1 、 T_2 、 T_3 、 T_4 、 T_5 、 T_6 、 T_7 、 T_8 等移動平均線變動天數的可調參數，包含 2 個買點子策略解(T_1 、 T_2)、(T_3 、 T_4)以及 2 個賣點子策略解(T_5 、 T_6)、(T_7 、 T_8)。買點決策準則為($SMA^{T_1} > SMA^{T_2}$) and ($SMA^{T_3} > SMA^{T_4}$)，賣點決策準則為($SMA^{T_5} < SMA^{T_6}$) and ($SMA^{T_7} < SMA^{T_8}$)。由於每支股票經由 n 次實驗(run)後，將會得到 n 個最佳解(global best)，而這 n 個最佳解均含有良好的交

易策略特質，並且可以拆解為各 $2n$ 個買賣子策略，以 (T_i, T_j) 表示。將以上具有良好交易特質子策略重複排列組合，可得各 $(2n)^4$ 個最佳解，稱為菁英候選解(elite candidate)，將這些菁英解經由目標適應函數運算(fitness function)後，並選取最好解作為此支股票之最佳交易策略。

步驟 6：選擇適合的分散式平台架構，建置巨量資料分析應用程式(服務)。

視資料分析需求選擇並架構適配之巨量資料分析運算平台。本研究以 EMC 框架【13】進行系統建構，並考量分析資料屬性(屬結構化、半結構化或非結構化資料)、分析資料儲存方式(SQL 或 NO-SQL)、應用分析結果回應服務方式(有無即時回應需求)、所應用資料分析方法的特性(適合併行運算或獨立完成運算)等因素。

步驟 7：對於資料子集合的資料屬性、可調參數及子集合本身等進行剖析，找出是否有樣態(pattern)存在。

四、結果與討論

4.1 實驗設計背景

本研究實驗是從台灣上市股票市場按類別區分為 33 類，除 5 類屬於憑證指數類別外，餘 28 類為選用樣本來源，每類股票依該類股按股價高中低(以實驗當日 2014/11/14 之收盤價為準)分三種，每類每群隨機各選一支作為樣本，因此總計 28 類、84 家支股票為訓練之樣本標的，從這些樣本資料標的中之每日收盤價作為資料來源，自動載入每日股票交易資料(含開盤價、收盤價、最高價、最低價、成交量等以作為訓練的資料來源。

每支股票完整實驗，分為訓練期(Training Period)與測試期(Testing Period)。訓練期自股票上市日期開始至 2011/12/31 止進行演化，每支股票執行 10 次訓練實驗(run)，每次訓練實驗(run)均執行 1,000 次演化，再從 10 次訓練實驗之最好的演化決策解。測試期自 2012/1/1 至 2014/11/14(實驗當日)為止，分別針對 EPSO、MMPSO、KennedyPSO 等策略，計算其測試期間之投資報酬率進行比較。

EPSO、MMPSO、KennedyPSO 演算法粒子個數參數為 50 個粒子($n=50$)，每個粒子以 X_i 表示之。每個粒子 X_i 內記憶二組資料，一是 8 個 SMA 變動天數位置(position)，分別以 X_{iT} 表示， X_{iT} 參數範圍為 6 至 400 天間， $1 \leq i \leq n$ ， $1 \leq T \leq 8$ 。另一組是對應 8 個 SMA 變動天數位置的速度參數(velocity)，以 V_{iT} 表示。亦即 V_{iT} 代表 X_{iT} 參數之移動方向與速度，其範圍設定為 -80 至 80 間， $1 \leq i \leq n$ ， $1 \leq T \leq 8$ 。 c_1 為粒子自信係數設定為 2(一般研究常用係數)， c_2 為粒子自信係數設定為 2(一般

研究常用係數)。 ω 為慣性權重值係數設定為 1.4，依實驗執行次數(runs)遞減至 0.4。股票投資期初資金(initial capital)為 1,000,000 元，每次交易至少買 50 張股票(每張股票 1000 股)，並且扣除交易成本。買入交易成本為交易金額的千分之 1.425，賣出之交易成本為交易金額的千分之 4.425。因考慮支付交易成本，每次買入以持有資金的 80% 購買股票。

4.2 股票交易策略績效探討

2012/1/1 至 2014/11/14 測試期間，EPSO 股票交易預測策略之 ROI 整體績效觀察得知，在實驗 84 支股票中，只有 1 支股票 ROI 績效為負，其餘 83 支股票 ROI 績效均為正數。其中 ROI 績效超過 50% 以上有 54 支股票(占全部測試股票之 64.29%)，超過 100% 以上有 31 支股票(占全部測試股票之 36.90%)，超過 200% 以上有 8 支股票(占全部測試股票之 9.52%)，而獲利表現超過 300% 以上有 3 支股票(占全部測試股票之 3.57%)，整體平均績效為 106.39%。

而 MMPSO 股票交易預測策略之 ROI 整體績效觀察得知，在實驗 84 支股票中，只有 25 支股票 ROI 績效為負，其餘 59 支股票 ROI 績效均為正數。其中 ROI 績效超過 50% 以上有 11 支股票(占全部測試股票之 13.10%)，超過 100% 以上有 4 支股票(占全部測試股票之 4.76%)，超過 200% 以上有 2 支股票(占全部測試股票之 2.38%)，而獲利表現超過 300% 以上有 1 支股票(占全部測試股票之 1.19%)，整體平均績效為 48.78%。

Kennedy PSO 股票交易預測策略之 ROI 整體績效觀察得知，在實驗 84 支股票中，只有 30 支股票 ROI 績效為負，其餘 54 支股票 ROI 績效均為正數。其中 ROI 績效超過 50% 以上有 6 支股票(占全部測試股票之 7.14%)，超過 100% 以上有 2 支股票(占全部測試股票之 2.38%)，超過 200% 以上有 2 支股票(占全部測試股票之 2.38%)，而獲利表現超過 300% 以上有 1 支股票(占全部測試股票之 1.19%)，整體平均績效為 13.13%。

Kennedy PSO 整體績效觀察得知，在實驗 84 支股票中，只有 24 支股票 ROI 績效為負，其餘 60 支股票 ROI 績效均為正數。其中 ROI 績效超過 50% 以上有 16 支股票(占全部測試股票之 33.33%)，超過 100% 以上有 7 支股票(占全部測試股票之 14.29%)，超過 200% 以上有 4 支股票(占全部測試股票之 5.95%)，而獲利表現超過 300% 以上有 1 支股票(占全部測試股票之 1.19%)，整體平均績效為 48.39%。

由表 1 可知 EPSO 整體 ROI 績效超過 100% 家數以上為 36.90%，遠超過 MMPSO(4.76%)及 Kennedy PSO(2.38%)；而超過 50% 家數以上為 64.29%，遠超過 MMPSO(13.10%)及 Kennedy

PSO(7.14%)。由上可得整體績效表現部分，以 EPSO 為最優，其次是 MMPSO、Kennedy PSO 次之，EPSO 績效排行前十大詳如圖 1 所示。

表 1：股票交易策略績效比較

ROI 績效獲利區間	EPSO 股票家數	MMPSO 股票家數	Kennedy PSO 股票家數
$ROI_{PSO} \geq 300\%$	3	1	1
$200\% \leq ROI_{PSO} < 300\%$	5	1	1
$100\% \leq ROI_{PSO} < 200\%$	23	2	0
$50\% \leq ROI_{PSO} < 100\%$	23	7	4
$0\% \leq ROI_{PSO} < 50\%$	29	48	48
$ROI_{PSO} < 0\%$	1	25	30

4.3 股票交易策略探討

本研究股票交易策略是由移動平均線的 SMA 天數比較決定，從 SMA 天數型態而看，SMA 天數多為長期天數(大於 120 天以上)，如表 2 所示，因此，本研究策略大多偏 SMA 長期天數移動平均線策略。

表 2：決策準則 SMA 天數型態表

交叉決策準則	Kennedy		
	EPSO	MMPSO	PSO
SMA 天數			
短期 SMA 天數個數	20	106	92
中期 SMA 天數個數	137	132	141
長期 SMA 天數個數	515	434	439

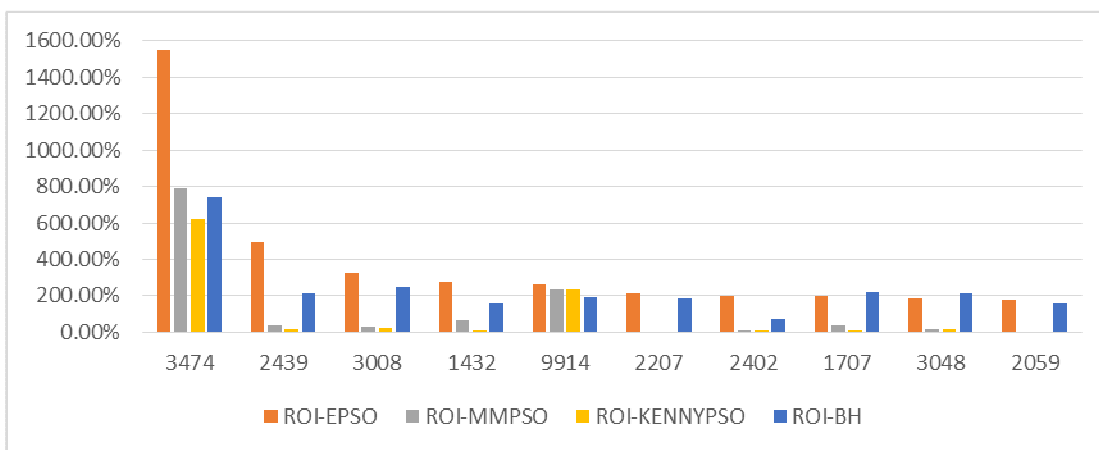


圖 1：EPSO 股票投資績效前十名

五、結論

本研究運用巨量資料分析資料觀點，結合知識發現方法，提出改良式粒子群最佳化演算法，透過股票交易資料、變動天數簡單移動平均線技術指標與其黃金死亡交叉決策準則，實現如何將股票知識、降低風險增加獲利等議題融入建構於巨量資料分析架構技術中，大幅提升股票投資報酬率，並且本研究策略的投資報酬率績效優於 MMPSO、Kennedy PSO 策略。

未來將利用資料採擷強大的資料分析與知識發現能力，針對股票交易預測策略準則態樣進行探索。另外，由於簡單移動平均線屬於落後指標，對於股票市場漲跌趨勢反應較為遲鈍，是否應再挑選一些領先指標輔助先行，或是挑選出高鑑別度之技術指標，以能分辨指標中雜訊及正確的資訊，將是未來預測模型精確的關鍵所在。

六、參考文獻

1. Atsalakis, George S., and Kimon P. Valavanis. "Surveying stock market forecasting techniques-Part II: Soft computing methods." Expert Systems with Applications, Vol. 36, No.3, pp. 5932-5941, 2009.
2. Sun, BaiQing, et al. "Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series." Neurocomputing, vol. 151, pp. 1528-1536, 2015.
3. Rostami, Neda, et al. "Comparison Models of Corporate Bankruptcy Prediction in Companies Listed in Tehran Stock Exchange by Particle Swarm Optimization Algorithm and Support Vector Machine Algorithm." Journal of American Science, Vol. 9, No.5, 2013.
4. Petchamé, Jordi, Àngela Nebot, and René Alquézar. "Quantitative and qualitative approaches for stock movement prediction." CCIA. 2012.

5. Boonpeng, Sabaitip, and Piyasak Jeatrakul. "Enhance the performance of neural networks for stock market prediction: An analytical study." Digital Information Management (ICDIM), 2014 Ninth International Conference on. IEEE, 2014.
6. Arik, Sercan, Sukru Burc Eryilmaz, and Adam Goldberg. "Supervised classification-based stock prediction and portfolio optimization." arXiv preprint arXiv:1406.0824, 2014.
7. Kwon, Yung-Keun, and Byung-Ro Moon. "A hybrid neurogenetic approach for stock forecasting." Neural Networks, IEEE Transactions on Vol.18, No.3, pp. 851-864, 2007.
8. Hsu, Ling-Yuan, et al. "Mutual funds trading strategy based on particle swarm optimization." Expert Systems with Applications, Vol.38, No.6, pp.7582-7602, 2011.
9. Begoli, Edmon, and James Horey. "Design principles for effective knowledge discovery from big data." Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), 2012 Joint Working IEEE/IFIP Conference on. IEEE, 2012.
10. Kennedy, James. "Particle swarm optimization." Encyclopedia of Machine Learning. Springer US, pp.760-766, 2010.
11. Gunasekarage, Abeyratna, and David M. Power. "The profitability of moving average trading rules in South Asian stock markets." Emerging Markets Review, Vol.2, No.1, pp.17-33, 2001.
12. Wu, Xindong, et al. "Data mining with big data." Knowledge and Data Engineering, IEEE Transactions on Vol.26, No.1, pp.97-107, 2014.
13. EMC Institute, VMware research team, Hu Jia Xi(translation)., Next step of Big Data: Big Data new strategy, technology and large-scale Web applications. Saatchi Times Press, 2014 .