國立政治大學金融學系

碩士學位論文

建構動態時間校正與聚類分析的選股模型實證研究

An Empirical Study of Stock Selection Strategies on Dynamic Time Warping and Cluster Analysis

指導教授：廖四郎 博士

研究生：曾子軒

中華民國一一一年六月

DOI:10.6814/NCCU202200715

摘要

本文探討動態時間校正及聚類分析應用於台灣股票市場之實證研究。目前相關的分類機器學習模型用於股票市場的文獻，大多以預測報酬率為基礎，輔以使用財務資料或技術指標而成，為了要將報酬率分類，必須定義標籤類別，使得我們必須主觀界定各個報酬率類別的邊界。再者，相似報酬率的股票對應的財務資料或技術指標，往往在時間上有領先或落後，導致在區分股票標的相似程度有困難，因此，本研究探討並試圖解決時間序列領先或落後的辨識相似度問題，並捨去需要事先定義類別標籤的方法，將相似的股票做聚類分析。

本文研究動態時間校正計算股票走勢相似度，並使用聚類分析將股票分群，從中建立交易策略發想，並比較不同的聚類分析模型所對應的結果。其結果顯示，動態時間校正更能有效辨識實際相似的股票走勢，其克服時間序列相似卻不同步的問題；聚類分析用於股票分群也有很好的表現，有助於選出報酬較好的標的。

關鍵詞：動態時間校正、聚類分析、Dynamic Time Warping、K-medoids、Fuzzy c-medoids

DOI:10.6814/NCCU202200715

ABSTRACT

This paper discusses the application of dynamic time warping algorithm and cluster analysis for stock selection and trading strategy in Taiwan stock market. Recently, the researches of classification machine learning model which are used in the stock market are mostly based on the classification of the predicted rate of return, supplemented by the use of financial data or technical indicators. Therefore, it is necessary to define different categories of rates of return. Furthermore, the financial data or technical indicators corresponding to stocks with similar returns usually lead or lag in time, which makes it difficult to distinguish the similarity of stocks. Therefore, we expect to solve the identification similarity of leading or lagging time series problems, and cluster the similar stocks. In this paper, dynamic time warping algorithm is used to calculate the level of similarity in the trend of stocks, and cluster analysis is used to group the stocks. And then, we build an empirical study of stock selection strategies, and compare the difference of clustering analysis models. The result shows that the dynamic time warping algorithm can more effectively identify the actual similar stock trends, and the cluster models we used also have good performance for stocks grouping, which help to select stocks with better returns.

Keywords：Dynamic Time Warping, Cluster Analysis, K-medoids, Fuzzy c-medoids

DOI:10.6814/NCCU202200715

目次

第一章 緒論.....1
第一節 研究動機與目的.....1
第二節 研究貢獻.....2
第二章 文獻回顧.....3
第一節 動態時間校正相關研究.....3
第二節 聚類分析相關研究.....4
第三章 研究方法.....6
第一節 距離度量.....6
第二節 動態時間校正.....6
第三節 聚類分析模型.....10
第四節 實證研究架構與績效衡量.....14
第四章 模型建構與實證分析.....20
第一節 資料來源與資料預處理.....20
第二節 模型設置及訓練.....22
第三節 實證結果.....25
第五章 結論與展望.....35
第一節 研究結論.....35
第二節 未來展望.....36
參考文獻.....37

DOI:10.6814/NCCU202200715

表次

表 1 K-mdeoids 分群績效比較 26

表 2 Fuzzy c-mdeoids 分群績效比較 29

表 3 模型間績效比較 33

DOI:10.6814/NCCU202200715

圖次

圖 1 校正路徑 7
圖 2 Sakoe & Chiba (1978)限制條件 8
圖 3 Sakoe & Chiba (1978)約束條件 9
圖 4 動態時間校正流程圖 15
圖 5 當沖策略流程圖 16
圖 6 K-medoids 聚類策略流程圖 17
圖 7 Fuzzy c-medoids 聚類策略流程圖 18
圖 8 使用歐式距離與不同群集數量之權益曲線 27
圖 9 使用曼哈頓距離與不同群集數量之權益曲線 28
圖 10 使用歐式距離及模糊性參數 2 與不同群集數量之權益曲線 30
圖 11 使用歐式距離及模糊性參數 3 與不同群集數量之權益曲線 31
圖 12 使用曼哈頓距離及模糊性參數 2 與不同群集數量之權益曲線 31
圖 13 使用曼哈頓距離及模糊性參數 3 與不同群集數量之權益曲線 32

DOI:10.6814/NCCU202200715

第一章 緒論

第一節 研究動機與目的

自從股票市場逐漸發展熱絡，許多投資的方法如春筍般地冒出，如基本面、技術面及籌碼面分析等，無論何種方法都有其各自的論述，同時也各有各自的支持者。近年來，隨著電腦運算速度漸漸提升，使得大數據與人工智慧的應用更加廣泛，也成為了現代熱門的研究領域之一。

將大數據結合金融領域的研究，從而形成了量化投資，有別於傳統的投資方法，量化投資運用大量的資料數據並搭配財務模型、數值方法、統計學等作為投資的研究方向，試圖在各種數據中找出相似的規律且有異常報酬的標的，相信同樣的邏輯會一再重複發生，進而形成投資決策，量化投資除了可以避免人為的情緒所造成錯誤的投資決策，也減少人工盯盤的時間，使我們能有更多時間進行資料研究和開發策略。

大數據的發展也造就了人工智慧與機器學習的一波熱潮，近年來也有不少文獻將其應用於金融的範疇形成另類的量化投資，例如機器學習和演算法用於選股，亦或是用於預測報酬率等，試圖讓模型學會成功的投資方法，創造超額的利潤。因此透過機器學習和大數據等方法用於量化投資，也成為了一個值得研究的範疇。

近年來，因各種國際事件及各國央行的貨幣政策，導致金融市場的波動度和成交量都有明顯的上升，這也引起我對於當沖的交易策略產生興趣，希望可以建構一套穩定獲利的策略，其中，股票 k 線型態學是許多人研究的方向，當投資人發現某些特定的型態出現時，預期未來會有異常的報酬產生，但不同的投資人對於型態的定義不盡相同，偏向於主觀的判斷，也因此有許多模糊的空間，無法大規模的統計與分類，難以形成一套標準化的交易策略。因此本文使用動態時間校

DOI:10.6814/NCCU202200715

正，試圖將股票走勢的相似度標準化，區分出不同的走勢形態；接著依據各個股票走勢的距離使用聚類分析將股票分群，分類出幾種型態的定義，並觀察各種型態後續的報酬。

第二節 研究貢獻

本研究顯示動態時間校正可以有效找出相似股票型態，並解決時間不同步的問題，對於辨識股票走勢相似程度有很大的幫助；聚類分析無論是 K-medoids 聚類或 Fuzzy c-medoids 聚類，用於股票分類皆有不錯的表現，有助於選出報酬較好的標的。

回顧文獻中，有不少文獻使用動態時間校正及聚類分析的相關研究，目前卻鮮少有文獻將兩者結合並用於金融領域，透過本文實證用於股票市場，可以得到不同走勢的股票未來對應的期望報酬確實有顯著的不同，也是一種獨特的選股方式。除此之外，金融大數據常常存在許多雜訊，一般較不容易透過分類機器學習模型得到顯著且有效的結果，因此本研究使用非監督式學習的機器學習模型—聚類分析，得到優於基準模型的表現，對於機器學習和金融領域的文獻做出了微薄的貢獻。

DOI:10.6814/NCCU202200715

第二章 文獻回顧

第一節 動態時間校正相關研究

在時間序列分析中，如果想要比較時間序列是否相似，一種簡單的方式即是在相同的時間下逐步比對是否相似，再綜合評斷是否為相似的序列，Keogh & Ratanamahatana (2005)中提到，許多文獻利用歐式距離或其衍生的距離度量衡量時間序列的相似性，然而這方法相當地脆弱且不夠彈性，因為不同的序列常常存在時間軸偏差，容易造成誤判，例如：每個人走路速度不一樣或說話速度不同，如果限制在相同的時間下進行比對，可能會得到不相似的結論，但比對的序列中可能是相似的，只是存在時間軸偏差的問題。

Berndt & Clifford (1994)中提到，人工辨識相似的時間序列圖形並不是件難事，但使用程式語言就變得很困難，困難點在於辨識圖形本身存在模糊性，造成辨識失敗率提高。一個好的辨識模型，例如應用於聲音辨識，我們應當著重於講者說的單子和句子，而不是說話的速度或音調的高低。使用動態時間校正計算序列的相似度可以有效解決時間軸偏差問題，是個強健且彈性的方法。

動態時間校正(Dynamic Time Warping; DTW) ，又稱動態時間扭曲，是一套衡量時間序列相似度的演算法，目前已廣泛應用於科學、醫學與財務領域，眾所周知的應用如語音辨識、圖形辨識和動作辨識等，事實上任何可以轉換為線性序列都可以使用 DTW 進行分析，也可用於部分形狀匹配應用。

DOI:10.6814/NCCU202200715

第二節 聚類分析相關研究

根據 DTW 計算時間序列的相似性之後，我們期望將相似的時間序列分類，我們可以用多種機器學習模型搭配 DTW 使用，本研究使用聚類分析作為 DTW 的搭配。聚類分析在機器學習中屬於非監督式學習的一種，意旨不用給定樣本點標籤(label)即可使用此模型，透過樣本點間的距離進行判斷從而產生群集的定義，Petitjean, Ketterlin, & Gançarski (2011)中提到，基於質心的聚類(centroid-based clustering)，如 K-means 聚類、層次聚類(hierarchical clustering)等為常見搭配 DTW 的方法，不過這類的聚類模型非常仰賴群集中心點的挑選，尤其用在時間序列的聚類上，隨著時間推移我們不容易找到適當的中心點，許多文獻嘗試為 DTW 定義中心點，但多呈現不夠準確或是干擾了聚類模型的收斂，因此更為適當的聚類模型搭配為 K-medoids 聚類。

K-medoids 聚類為 Kaufman & Rousseeuw (1990)提出，是屬於硬聚類的一種，代表每個樣本點只會被歸類至其中一個類別，沒有模糊的空間。K-medoids 聚類和 K-means 聚類類似，皆是聚類分析的一種，Jin & Han (2010)、Labroche (2010)中提到，相較於 K-means 聚類，K-medoids 聚類面對極端值與離群值有更好的辨識力，兩個模型皆需要設定集群的中心點，不同的方式在於 K-means 聚類計算每個集群內樣本點的質心；K-mdeoids 聚類則是從樣本點中選擇，挑選一個最具代表性的點作為集群的中心點。

除了 K-medoids 聚類外，本文亦同時比較 Fuzzy c-medoids 聚類，Fuzzy c-medoids 聚類屬於模糊聚類的一種，也有人稱之軟聚類，是由 Krishnapuram et al. (2001)所提出，有別於 K-medoids 聚類，Fuzzy c-medoids 聚類將樣本點的分類轉為機率，意旨假設我們要將樣本點分類成 k 個類別，經過計算每個樣本點將會得到 k 個機率，且機率加總等於 1，這表示在 Fuzzy c-medoids 聚類中，每個樣本點

DOI:10.6814/NCCU202200715

並非絕對屬於哪一個類別，而是根據樣本點與各個中心點的距離，計算該樣本點於屬於每一個群集的機率，Fuzzy c-medoids 聚類保留了距離的資訊，對於某些和多個中心點距離相似的模糊樣本點，K-medoids 聚類選擇之中最近的中心點並給予其類別；而 Fuzzy c-medoids 聚類則是以多個相似的機率值代表之。

DOI:10.6814/NCCU202200715

第三章 研究方法

第一節 距離度量

給定維度為 n 的兩個點$\{x_i, i = 1,2, \dots, n\}$，$\{y_i, i = 1,2, \dots, n\}$，我們定義$d(x, y)$代表 x 與 y 之間的距離，本研究嘗試兩種距離度量，分別為歐氏與曼哈頓距離，詳細說明如下：

歐幾里得距離距離(Euclidean distance)，又稱歐式距離，為 x 與 y 之間的直線距離，公式如下：

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

曼哈頓距離(Manhattan Distance)為 x 與 y 之間每一個維度的距離加總，公式如下：

$$d(x,y) = \sum_{i=1}^{n}|x_i - y_i|$$

第二節 動態時間校正

假設有兩個時間序列，其中 x 的長度為 m，以$\{x_i, i = 1,2, \dots, m\}$表示，y 的長度為 n，以$\{y_i, i = 1,2, \dots, n\}$表示，且 x 和 y 皆是維度為 Q 的多變量時間序列，動態時間校正計算 x 與 y 的距離，試圖解決兩條時間序列可能存在時間軸偏差的問題，並找到一組最佳時間對應路徑用來衡量 x 與 y 的相似性，我們稱此路徑為校正路徑(warping path)或稱為校正函數(warping function)，如以下公式：

$$F = c_1, c_2, \dots, c_k, \dots, c_K \ ， \ c_k = (i_k, j_k) \ ， \ k = 1,2, \dots, K$$

DOI:10.6814/NCCU202200715

其中 c_i 代表一組座標，代表每個 x 與 y 的對應點，即點對點的路徑，如圖 1 為例。

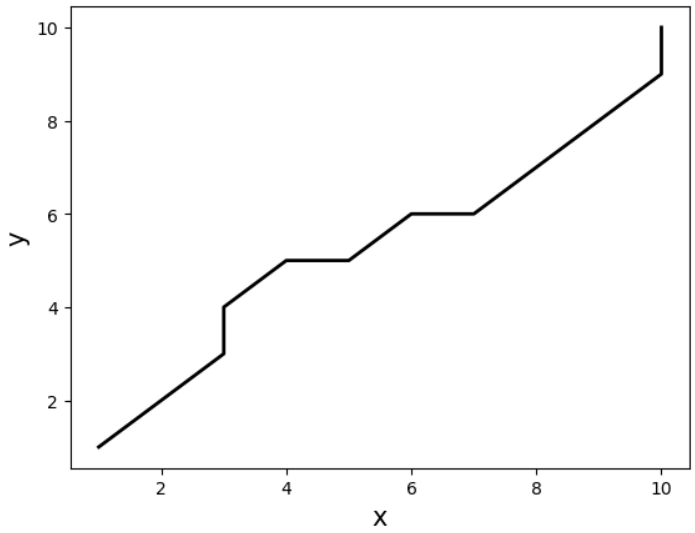

圖 1 校正路徑

在校正路徑中，我們給定四個基本的限制：

1. 起點與終點限制

$$c_1 = (1,1) \quad , \quad c_K = (m,n)$$

x 與 y 的起始點必須相連接，同樣地，x 與 y 的終點也必須相連接，這意味著端點為「頭對頭、尾對尾」的比對。

2.單點或多點匹配

x 與 y 的每一個點都必須相互配對，不存在任意一點沒有配對，且每一個點容許可以與多個點匹配。

3.單調條件(monotonic condition)

$$i_k - i_{k-1} \geq 0 \quad , j_k - j_{k-1} \geq 0$$

DOI:10.6814/NCCU202200715

校正路徑會隨時推移向前，也就是c_k的點相較於c_{k-1}只會增加或和其相同，並不會向後搜尋，我們稱此限制為單調條件或單調遞增條件。

4.附近連續條件(local continuity condition)

在尋找校正路徑時，我們會向時間序列附近的點搜尋，尋找最佳的校正路徑，不過也不會搜尋到太遠的點，因此我們會限制搜尋的範圍並稱之為附近連續條件，附近連續條件有多種形式，例如：Sakoe & Chiba (1978)提出的限制條件，公式如下：

$$i_k - i_{k-1} \leq 1 \quad , j_k - j_{k-1} \leq 1$$

假設校正路徑中的第 k 個點$c_k = (\mathrm{i}, \mathrm{j})$，那麼$c_{k-1}$則可能為$(\mathrm{i}-1, \mathrm{j})$，$(\mathrm{i}, \mathrm{j}-1)$，$(\mathrm{i}-1, \mathrm{j}-1)$，只搜尋現在或上一點，如圖 2 所示。

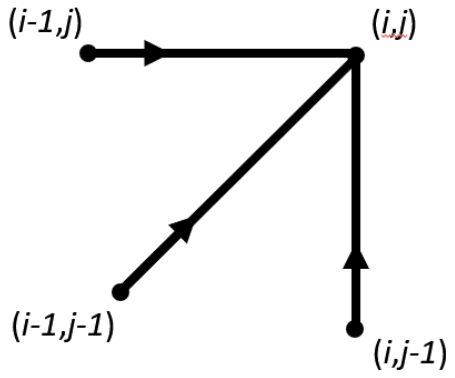

圖 2 Sakoe & Chiba (1978)限制條件

加上附近連續性的限制條線，可以減少計算成本以提高動態時間校正演算法的效率，但缺點在於 x 與 y 時間序列如果時間軸偏差較大，則無法找到最佳的校正路徑。我們又參考了 Sakoe & Chiba (1978)文中提出另一種的約束條件：

$$i_k - i_{k-1} \leq n \quad , j_k - j_{k-1} \leq n$$

DOI:10.6814/NCCU202200715

則c_{k-1}可能為{i, i − 1, … , i − n}與{j, j − 1, … , j − n}的組合，代表我們在考慮校正路徑時只會參考現在或前 n 範圍內的點，如圖 3 所示。

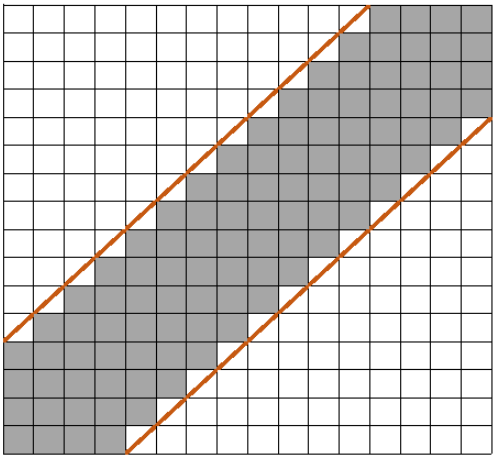

圖 3 Sakoe & Chiba (1978)約束條件

此約束條件也同樣可以提升演算法的效率，同時又考慮時間軸偏差較大的問題，試圖在兩者之中取得平衡。

接著說明動態時間校正的流程，設定 $\{x_i, \mathrm{i} = 1,2, \dots, \mathrm{m}\}$ 與 $\{y_j, \mathrm{j} = 1,2, \dots, \mathrm{n}\}$ 兩時間序列，其動態時間校正距離為 $\mathrm{D(i,j)}$ ，我們將計算所有的 $\mathrm{D(i,j)}$，$\mathrm{i} = 1,2..\mathrm{m}$、$\mathrm{j} = 1,2, \dots, \mathrm{n}$。

一開始我們先定義任意兩點x_i, y_j的距離為：

$$d_{ij} = d(x_i, y_j) \text{ ，} i = 1,2, \dots, m \text{ 、} j = 1,2, \dots, n$$

距離度量使用我們先前決定的度量，例如：歐式距離，接著我們即可開始計算動態時間校正距離$D(i,j)$，其中邊界值給定為：

$$D(0,0) = 0$$

$$D(0,j) = \infty \quad \forall j = 1,2,3, \dots, n$$

DOI:10.6814/NCCU202200715

$$D(i,0) = \infty \quad \forall i = 1,2,3,\dots,m$$

如此設定讓我們在初始點得以選擇 x 與 y 的起始點，即$c_1 = (1,1)$，其餘的動態時間校正距離為：

$$D(i,j) = d_{ij} + min\{D(i-1,j), D(i,j-1), D(i-1,j-1)\}$$

$$\forall i = 1,2,3,\dots,m \text{ 、} j = 1,2,3,\dots,n$$

重複上式的遞迴關係，當我們計算至終點 $c_K(m,n)$ 時，終點的動態時間校正距離D(m, n)即求得x, y的時間序列校正距離。

第三節 聚類分析模型

一、K-medoids 聚類模型

假設有 n 個樣本點，我們想要將這 n 個樣本點分成 K 個類別(n > K)，為了衡量樣本之間的相似性，我們需要計算所有樣本之間的距離，我們需要先定義距離度量，一般使用歐式距離或曼哈頓距離作為距離度量，我們也可以使用動態時間校正距離作為距離度量。

以下是 K-medoids 聚類的詳細步驟，我們將拆分為三個步驟演示：

1. 選擇中心點初始值

值得注意的是初始化中心點的方法可以有很多種，本文採用 Krishnapuram et al. (2001)提出的方法，從 n 個樣本隨機選取 K 個點作為集群中心的初始值

$$\mu_c^{(0)} \in R^d, c = 1,2,\dots,K$$

DOI:10.6814/NCCU202200715

$$iter = 0$$

其中R^d為全部的樣本點，c 代表集群的編號，iter 為迭代次數。每個樣本點都尋找其對應最近的 $\mu_c^{(0)}$，作為初始的群集的類別：

$$S_c^{(0)} = \left\{x_i: \left\|x_i - \mu_c^{(0)}\right\| \leq \left\|x_i - \mu_{c^*}^{(0)}\right\|, \forall i = 1, \ldots, n\right\}$$

由於金融數據存在許多雜訊，在模型訓練的過程容易因不同的起始點而有不同的訓練結果，因此我們重複執行該演算法多次，每次選擇不同的 K 個 $\mu_c^{(0)}$ ，如此可以有效訓練出較好的模型結果，也可以避免過擬合或陷入局部最佳解。

2. 迭代中心點

隨機選擇新的中心點 $\mu_c^{(t+1)}$ 作為群集的中心點，每個樣本點同樣尋找其對應最近的 $\mu_c^{(t+1)}$，作為群集的類別

$$S_c^{(t+1)} = \left\{x_i: \left\|x_i - \mu_c^{(t+1)}\right\| \leq \left\|x_i - \mu_{c^*}^{(t+1)}\right\|, \forall i = 1, \ldots, n\right\}$$

$$iter = iter + 1$$

如果群集內樣本點到 $\mu_c^{(t+1)}$ 的距離加總小於舊的中心點 $\mu_c^{(t)}$ 距離加總，則用 $\mu_c^{(t)}$ 迭代 $\mu_c^{(t-1)}$ 。

3. 演算法終止條件

在執行步驟 2 時，如果中心點有進行迭代，則重複執行步驟 2；反之如果新舊中心點一樣，則停止演算法，得到最終的 $\mu_c^{(t)}$。

$$S_c^{(t+1)} = S_c^{(t)} \ or \ iter = MAX\_ITER$$

DOI:10.6814/NCCU202200715

二、Fuzzy c-medoids 聚類模型

假設同樣是 n 個樣本，想要將這 n 個樣本點分成 K 個類別(n > K)，為了衡量樣本之間的相似性，我們需要計算所有樣本之間的距離，我們可以使用 DTW 作為距離度量，在此我們定義兩個樣本點x, y的距離為d(x, y)。

以下是 Fuzzy c-medoids 聚類的詳細說明：

1. 選擇中心點初始值

同樣採用(Krishnapuram et al. (2001)提出的方法，從 n 個樣本隨機選取 K 個點作為集群中心的初始值

$$\mu_c^{(0)} \in R^d, c = 1,2, \dots, K$$

$$iter = 0$$

接著定義u_{ij}為：

$$u_{ij} = \frac{(1/d(x_j, c_i))^{1/(m-1)}}{\sum_{k=1}^{c}(1/d(x_j, c_k))^{1/(m-1)}}，i = 1, \dots, K，j = 1, \dots, n$$

其中d(x, c)表示樣本點到中心點的距離，m 為模糊性參數。u_{ij}的分母為樣本點到所有中心點的距離總和，分子為樣本點到其中一個中心點的距離，大致上可以看成與d(x, c)成反比，當樣本點距離某一個中心點相近時，u_{ij}值越大。

2. 迭代中心點

接下來我們定義目標函數，求解最小化目標函數：

DOI:10.6814/NCCU202200715

$$q = arg\,min J(k) = \sum_{i=1}^{K}\sum_{j=1}^{N} {u_{ij}}^{m} \cdot \left\| x_j - c_i \right\|^2$$

3. 演算法終止條件

在執行步驟 2 時，如果中心點有進行迭代，則重複執行步驟 2；反之如果新舊中心點一樣，則停止演算法，得到最終的 $\mu_c^{(t)}$。

$$S_c^{(t+1)} = S_c^{(t)} \;\; or \;\; iter = MAX\_ITER$$

DOI:10.6814/NCCU202200715

第四節 實證研究架構與績效衡量

一、實證研究架構

本研究結合動態時間校正及聚類分析模型，建構一個當沖的交易策略。其中動態時間校正用來計算股價報酬時間序列的距離，接著傳入聚類分析查看分群的結果。

為了避免過擬合和的現象發生，本文事先以日期作為訓練集資料和測試集資料的分隔，將台股 9:00~11:00 一分鐘 k 棒的報酬時間序列作為動態時間校正的輸入值，計算不同時間或標的時間序列相似性，因此可以得到一個 N*N 的距離矩陣，其中 N 代表所有樣本的個數，矩陣中的每一個數值代表兩條時間序列的動態時間校正距離，藉此可以分析各個報酬時間序列相似的程度；之後再將距離矩陣輸入至聚類模型分群，在此我們分別使用聚類分析中的硬聚類—K-medoids 聚類及軟聚類—Fuzzy c-medoids 聚類，其中 K-medoids 聚類的分群結果明確，會得到每一條序列所屬的群集，沒有任何的模糊空間；Fuzzy c-medoids 聚類的分群結果則是告訴我們每一條序列分別屬於每一個群集的機率，讓我們可以自行操作後續的策略，流程圖如圖 4 所示。

DOI:10.6814/NCCU202200715

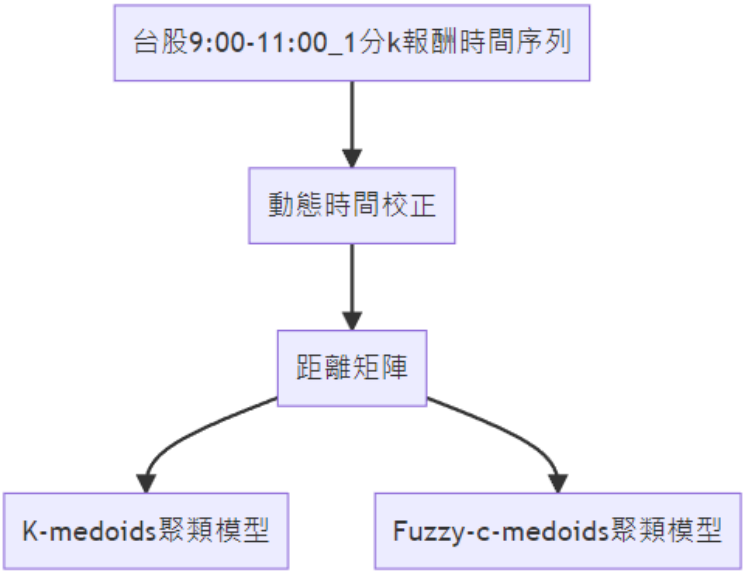

圖 4 動態時間校正流程圖

接著本文對所有的股票實行相同的策略，所有的股票皆為當日沖銷並不會留倉至隔天，由於大多數的股票在盤中時多呈現開高走低的盤勢，因此將策略訂定為放空策略，之後經過流動性篩選、停損停利等流程，計算每一檔股票的當沖報酬率，其中，流動性篩選是為了找到可以交易的股票，如果該股票僅有少數的成交量，並不適合做為當日沖銷的標的，本文使用過去幾天的平均周轉率作為流動性指標的參考，選出過去幾天平均周轉率較高的股票作為可以交易的標的；進場價格限制則是限制進場時，價格不能為接近漲停價或跌停價，以免造成收盤無法回補，或進場放空不久後就跌停，報酬率無法填補手續費等問題；停損停利的方式則是固定停損及停利並加上漲停前強制回補，確保風險可以獲得一定的控制，流程圖如圖 5 所示。

DOI:10.6814/NCCU202200715

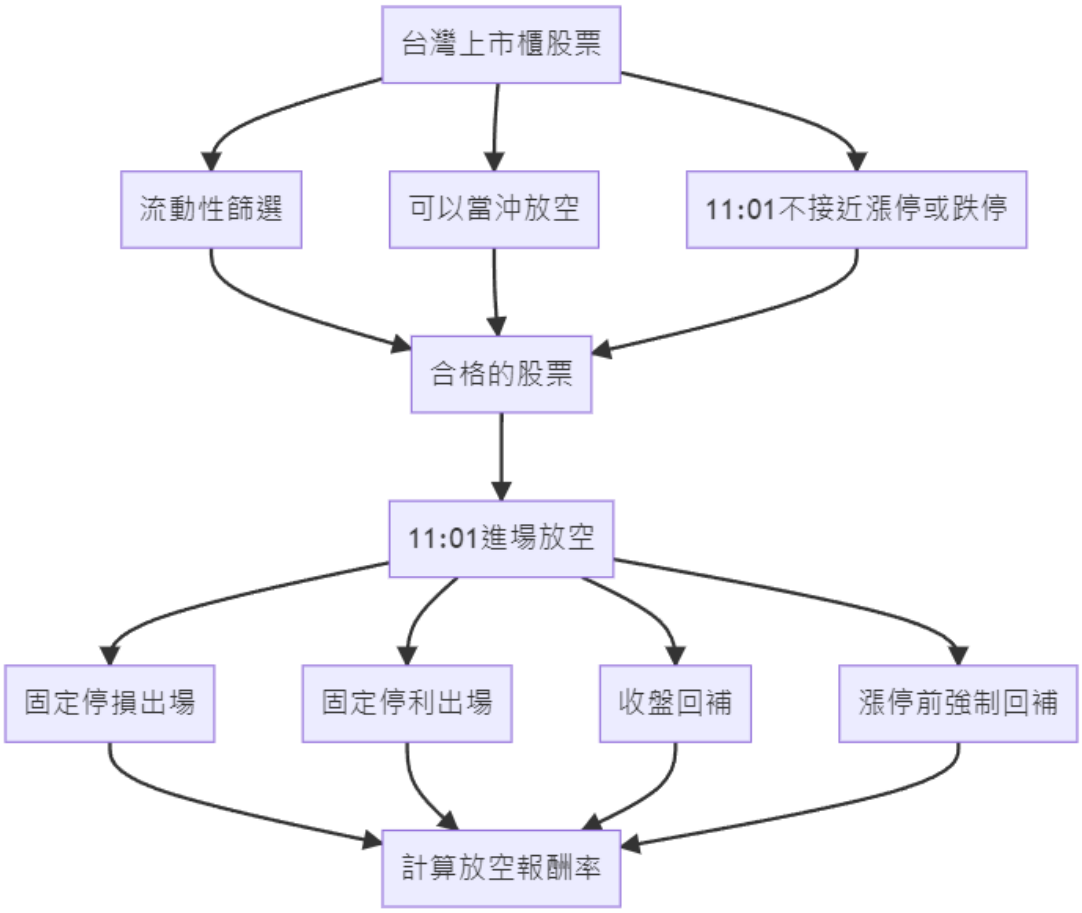

圖 5 當沖策略流程圖

接著我們說明聚類模型的策略流程，K-medoids 聚類和 Fuzzy c-medoids 聚類在挑選合格股票的方式有些不同，其中 K-medoids 聚類會透過距離矩陣將每一檔股票進行分類，我們可以得到每一檔股票分別屬於的群集，接著計算每一群的平均報酬率，挑選平均報酬率較高的群集作為合格的群集，之後測試集資料倘若有股票 9:00~11:00 的報酬時間序列被分類到合格的群集，我們就執行圖 4 的策略，流程圖如圖 6 所示。

DOI:10.6814/NCCU202200715

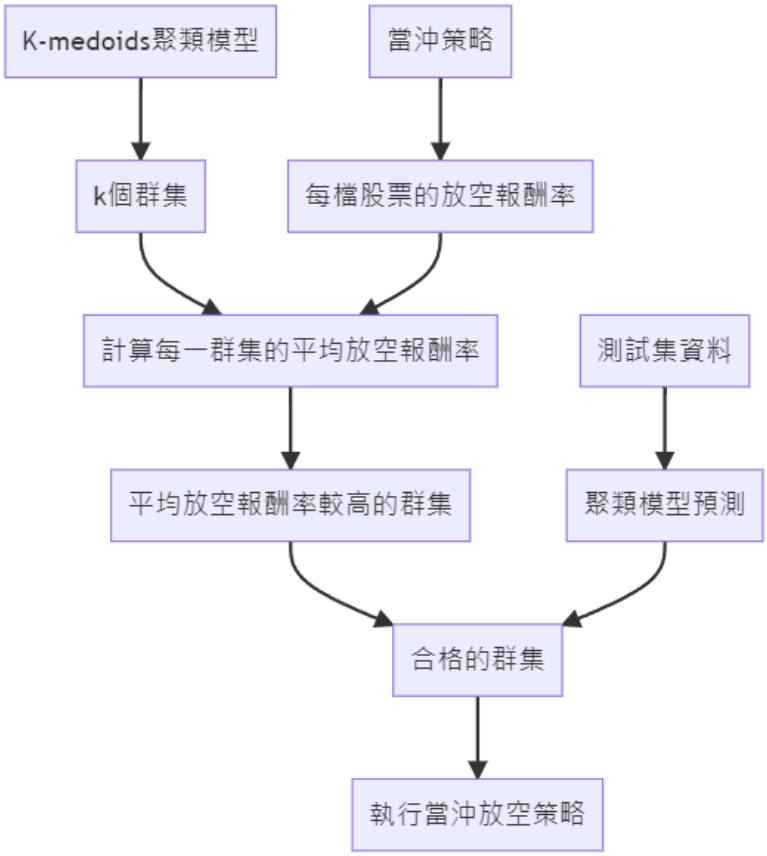

圖 6 K-medoids 聚類策略流程圖

Fuzzy c-medoids 聚類則是透過距離矩陣將每一檔股票進行分類，我們可以得到每一檔股票分別屬於這 k 個群集的機率，透過當沖策略算出每一檔股票的報酬率乘以它被分到各個群集的機率，得到每一群集的平均報酬率，之後.透過測試集資料的股票 9:00~11:00 報酬時間序列傳入軟聚類模型中，同樣會得到該樣本屬於每一群的機率，並將該機率乘以每一群的平均報酬率得到該股票的期望報酬率，如果期望報酬率夠高，則執行圖 4 的策略，流程圖如圖 7 所示。

DOI:10.6814/NCCU202200715

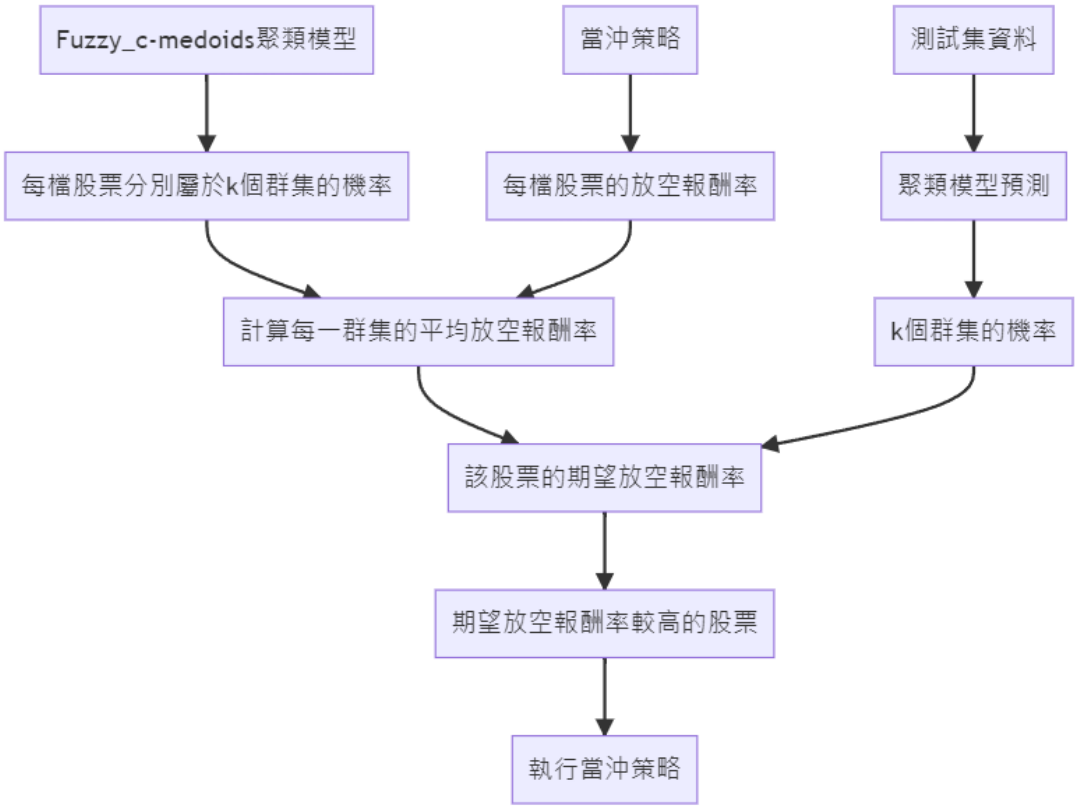

圖 7 Fuzzy c-medoids 聚類策略流程圖

二、績效衡量

我們使用多種績效衡量指試圖在報酬及風險中取得平衡，其中包含總報酬率、手續費以及一些衡量投資績效的指標，例如：夏普比率、索提諾比率、最大回撤等，績效指標詳細的計算方式如下：

$$Sharpe\ Ratio\ =\ \frac{Mean(re)}{[sd(re)*\sqrt{252}]}$$

$$Sortino\ Ratio\ =\ \frac{Mean(re)}{[Mean(nr\hat{}2)*\sqrt{252}]}$$

$$Maximum\ Drawdown\ =\ max\{cummax[cumsum(re)]\ -\ cumsum(re)\}$$

其中上述三條式中的 Re 為每日的報酬率，我們假設一年的交易日為 252 天，

DOI:10.6814/NCCU202200715

夏普比率及索提諾比率我們假設無風險利率為 0，夏普比率展示投資人面對單位風險時可以獲得的報酬；索提諾比率中，nr 為下方的報酬率，即虧損的報酬率，當報酬為正時則用 0 取代之，索提諾比率的分母透過計算下方風險的波動度，以此衡量投資人面對一單位下方風險可以獲得的報酬;最大回撤說明在過去的回測當中,投資人的資產從最高點回落的最大值,預期未來可能發生類似的虧損幅度。

交易稅的部份我們參考國泰證券的手續費，買賣各為成交金額的0.001425 * 0.28；證交稅當日沖銷減半，為成交金額的 0.0015。

三、基準模型

除了報酬及風險的衡量，我們更加入了大盤和基準策略的比較，其中包含加權報酬指數、台灣元大 50 ETF 以及當沖放空的基準策略，共三個基準策略進行比較，加權報酬指數為台灣加權指數加上所有的配息，所組成加權還原指數，並買進且持有的報酬率形成加權報酬指數；台灣元大 50 ETF，股票代號 0050，為元大發行的指數型基金，持股為台灣前 50 大市值的上市公司；當沖放空基準策略如圖 4 所示，不經過聚類模型的選股操作。

DOI:10.6814/NCCU202200715

第四章 模型建構與實證分析

第一節 資料來源與資料預處理

一、資料來源

本文模型使用資料從台灣證券交易所選取，日期區間為 2020 年 4 月至 2022 年 3 月，包含台股所有上市櫃股票的日內成交回報並事先將其轉換成 1 分鐘的頻率；以及從 CMoney 理財寶選取同樣期間和範圍台股的每日股價、週轉率、開盤參考價、漲停參考價、跌停參考價、可當沖做多股票、可當沖放空股票等資料；基準策略的資料為加權報酬指數及台灣 50 ETF 股價資料，兩者皆從 TEJ 台灣經濟新報資料庫取得。

二、資料預處理

因為大量資料較不利於分析，因此本文對上述資料進行一系列預處理，除了將資料簡化之外，使用 CMoney 的資料將實證研究更貼合實際狀況。本文實作的資料處理包含：可當沖標的篩選、流動性篩選、空值處理、資料標準化、切分樣本內外等，詳細說明如下：

1. 可當沖標的篩選

首先，因本文研究當沖放空的交易策略，須符合台灣證券交易所可當沖的法規限制，因此利用 CMoney 每日盤前發布的可當沖放空標的，選出可以當沖放空的股票。

DOI:10.6814/NCCU202200715

2. 流動性篩選

台灣上市櫃公司有不少股票每日交易量較少，導致想要交易這些股票時會有很大的滑價或是無法成交，因此本文從台灣所有的股票中先篩選流動性。篩選流動性有許多方法，參考他人的做法，篩選每一檔股票過去五天的平均日週轉率大於 5%，作為股票流動性的篩選。

3. 空值處理

由於某些股票並沒有在每一分鐘都有成交，基於資料的完整性，將沒有成交的股價填入上一分鐘的收盤價。CMoney 可當沖做多股票及可當沖做空股票也有一些空值，基於不知道這些股票是否可以當沖，我們假設這些股票不能當沖，並排除這些股票。

4. 資料標準化

將每一分鐘股價資料和該股票當天第一分鐘收盤價相除，以報酬率替代股價時間序列，以解決股價比例不同的問題。

5. 切分樣本內外

為了避免過度擬合，本文切分樣本內及樣本外，並僅使用樣本內的資料進行訓練，樣本內的時間區間為 2020 年 4 月至 2021 年 1 月，共 10916 個樣本點，每個樣本點為 9:00~11:00 之 1 分鐘 k 報酬時間序列；樣本外的時間為 2021 年 2 月至 2022 年 3 月，共有 19466 個樣本，每個樣本點同樣為 9:00~11:00 之 1 分鐘 k 報酬時間序列；全樣本總共有 30382 調時間序列。

DOI:10.6814/NCCU202200715

第二節 模型設置及訓練

一、動態時間校正模型設置

動態時間校正需要設定的參數包含：距離度量、約束條件的寬度。由於傳統的動態時間校正須計算兩兩時間序列的距離藉此比對相似度，但也因此產生大量的計算成本，以兩條長度分別為 n,m 的時間序列為例，其運算複雜度為 O(nm)，而本文使用的 9:00~11:00 之 1 分鐘 k 棒長度為 120，因此運算複雜度為 O(14400)，再加上 3 萬多個樣本點需要計算，要計算完全部的時間序列距離並不切實際，因此我們使用 Sakoe-Chiba(1978)提出改良的約束方法，如下烈式所示：

$$i_k - i_{k-1} <= n \ , j_k - j_{k-1} <= n$$

其中 n 為我們需要決定的參數之一，也有人稱之上下界約束(window size)，透過訂定上下界約束，藉此降低運算複雜度，大幅提升運算速度，根據 Keogh & Ratanamahatana (2005)文獻中建議，將上下搜尋的範圍(window size)設定為 n/10，其中 n 為時間序列長度，本文使用 9:00~11:00 的 1 分鐘 k 的報酬時間序列，時間序列長度為 120，按照文獻應將搜尋範圍設定為 12，但考量 1 分鐘 k 的金融資料容易含有較大的雜訊，因此給予較寬鬆的限制，將搜尋的上下界範圍設定為 20；第二個參數為距離度量，我們選擇曼哈頓距離及歐氏距離並進行比較。

DOI:10.6814/NCCU202200715

二、聚類分析模型設置

1. K-medoids 聚類模型設置

K-medoids 聚類需要設定的參數包含：需要分群的數量 k、最大中心點迭代次數 iter.max、每個模型重複訓練次數 nrep 以及距離度量。

關於分群的數量 k，每一個群集代表一種不同的報酬時間序列走勢，如果 k 的值太小，代表分群的數量較少，較不容易有過擬合的狀況，但也有可能使得不太相似的報酬時間序列因此被分至同個群集，因此嘗試多個不同的 k 值試圖在過擬合和分群有效性取得平衡，本文嘗試測試的 k 值有 50,100,200,500 個群集，比較不同的群集數量並找出最佳解。

最大中心點迭代次數 iter.max，模型的一開始選定 k 個點當作每一個群集的中心點，並計算每個樣本點到各個中心點的距離，試圖找到同群集的樣本距離中心點和最小的點，由於金融數據有許多的雜訊，即使經過多次新的中心點迭代也很容易得到局部最佳解，並耗費大量的運算力和時間，因此本文將最大中心點迭代次數 iter.max 設定為 10，即最多迭代 10 次，如果迭代後的中心點不變則提早結束模型訓練。

設定每個模型重複訓練次數 nrep，模型的一開始會隨機選定 k 個樣本點當作每一個群集的中心點，由於在模型訓練的過程很容易掉入局部最佳解，初始值是訓練出好模型重要的因素，但我們並不知道每次選擇初始值是否能為我們訓練出優良的結果，因此重複且隨機選用多次不同的迭代點可以有效解決這個問題，本文設置 nrep 為 20，意旨每個群集重複訓練模型模型 20 次，且每次隨機選擇不同的初始值。

距離度量我們使用動態時間校正，並分別搭配歐氏距離和曼哈頓距離並比較

DOI:10.6814/NCCU202200715

兩者的結果。

2. Fuzzy c-medoids 聚類模型設置

Fuzzy c-medoids 聚類需要設定的參數包含：軟聚類的模糊性 m、聚類分析需要分群的數量 k、最大中心點迭代次數 iter.max、以及距離度量。

軟聚類的模糊性參數 m 為 Fuzzy c-medoids 聚類重要的參數，一般會將此參數設定為大於 1 的值，Labroche (2010) 中提到，當模糊性參數 m 愈接近 1 時，該結果會和 K-medoids 聚類結果相似；隨著 m 愈來愈大時，樣本點與中心點的距離因素會被加重考慮，模糊性會增加並且質心集中在數據的中心，本文嘗試將 m 設定為 2。

分群的數量 k 我們同樣嘗試測試 50,100,200,500 個群集，比較不同的群集數量並找出最佳解；最大中心點迭代次數 iter.max 本文同樣設定為 10，即最多迭代 10 次，如果迭代後的中心點不變則提早結束模型訓練；距離度量我們使用動態時間校正，並分別搭配歐氏距離和曼哈頓距離並比較兩者的結果。

DOI:10.6814/NCCU202200715

第三節 實證結果

本研究使用 2021 年 2 月至 2022 年 3 月台股資料作為樣本外的研究資料，除了各個模型、參數間的比較，我們也加入了基準模型的比較。本節共分為三個部分，第一、二部分將會介紹本研究使用的兩種模型，比較相同模型、不同參數，並同時和基準模型進行對比，第三部分將會比較模型之間的不同。

一、K-mdeoids 聚類結果

我們比較 K-mdeoids 聚類，根據不同的分群數量及距離度量所得到的績效圖，其中分群數量選擇 50、100、200、500 等四種，距離度量使用歐式距離及曼哈頓距離。

從表 1 可以看出，透過 K-mdeoids 聚類分群之後，夏普比率皆大於 3，索提諾比率皆大於 5，最大回撤皆控制在 4%至 5%，皆顯著優於沒有分群的基準模型，同時也優於台灣元大 50 ETF 與台灣加權報酬指數，因次可以說明透過 K-mdeoids 聚類用於台灣股票報酬時間序列分群，並篩選特定的群集是有效的，接著我們將會進行模型內不同參數的比較。

DOI:10.6814/NCCU202200715

表 1 K-mdeoids 分群績效比較

集群數量	距離度量	年化報酬率	夏普比率	索提諾比率	最大回撤
50	Euclidean	0.654	3.071	5.1721	0.0575
	Manhattan	0.6368	3.161	5.3488	0.0607
100	Euclidean	0.704	3.5228	6.1245	0.0475
	Manhattan	0.7212	3.4607	6.0092	0.051
200	Euclidean	0.7961	4.0512	7.5423	0.0431
	Manhattan	0.6804	3.5817	6.4962	0.0454
500	Euclidean	0.582	3.1469	5.4661	0.0519
	Manhattan	0.5765	3.0592	5.2438	0.0554
基準模型		0.339	1.9591	3.4041	0.061
0050 ETF		0.0725	0.3993	0.5865	0.1236
台灣加權報酬指數		0.1433	0.8258	1.1821	0.1335

DOI:10.6814/NCCU202200715

如圖 8 及圖 9 中所示，可以觀察到不同的分群數量隨著時間的推移，績效有越來越顯著的差異，其中又以 100,200 個群集數量表現較好，無論是總報酬率、各種績效比率都表現較為突出，由此可知在本研究的架構上，這個範圍的分群數量有較好的表現，我們推測當分群數量設定為 50 時，因為分群數量較少，許多較不相似的報酬時間序列會被分至同一個群集，造成每個群集的雜訊增加；相對的，500 個群集則將報酬時間序列分類太多個類別，許多群集內的樣本數不足為無效的群集，但篩選表現較好的群集時依然被我們納入，造成績效下降。至於距離度量，我們分別使用歐式距離及曼哈頓距離，兩種距離度量在本實驗的結果中並沒有明顯的差異。

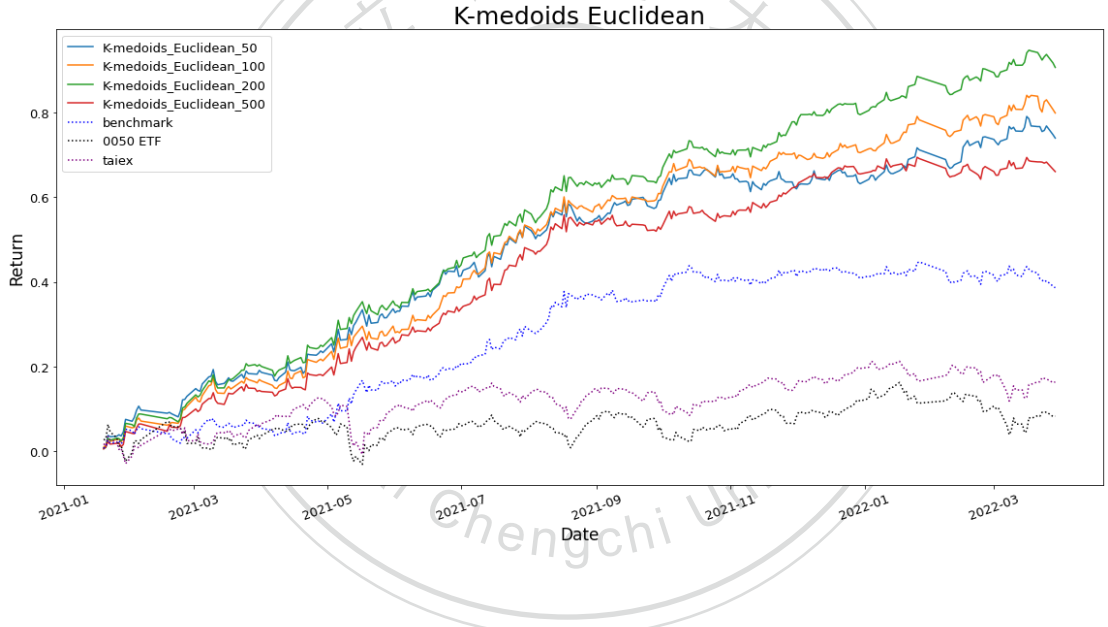

圖 8 使用歐式距離與不同群集數量之權益曲線

DOI:10.6814/NCCU202200715

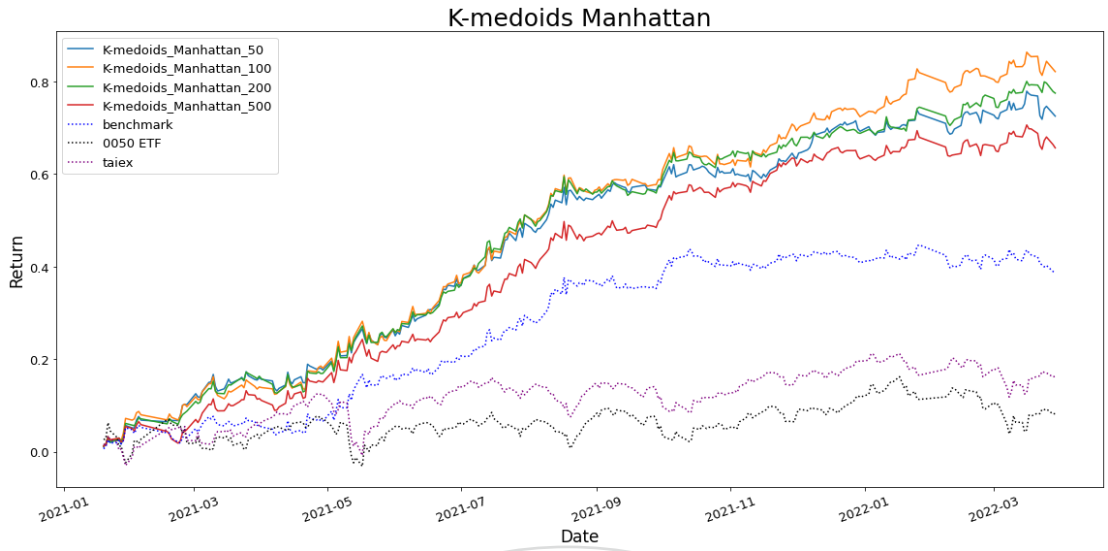

圖 9 使用曼哈頓距離與不同群集數量之權益曲線

二、Fuzzy c-mdeoids 聚類結果

Fuzzy c-mdeoids 聚類我們同樣使用 50、100、200、500 等四種分群數量，歐式距離及曼哈頓距離等兩種距離度量，模糊度我們分別測試 2、3 等兩種參數。

從表 2 可以看出，透過 Fuzzy c-mdeoids 聚類分群之後，夏普比率普遍也都有 2 以上，索提諾比率則介於 3 至 5，最大回撤則沒有控制得很好，幾乎都超過 10%，只有部分的分群結果優於基準模型，不過大部分仍舊優於台灣 50 ETF 與台灣加權報酬指數，說明 Fuzzy c-mdeoids 聚類將台灣股票報酬時間序列分群，並用群集機率和群及預期報酬計算個股預期報酬是部分有效的，接著我們將會進行模型內不同參數的比較。

DOI:10.6814/NCCU202200715

表 2 Fuzzy c-mdeoids 分群績效比較

集群數量	距離度量	模糊性參數	年化報酬率	夏普比率	索提諾比率	最大回撤
50	Euclidean	2	0.5648	2.2382	3.5314	0.2122
	Manhattan	2	0.427	1.5441	2.309	0.2403
100	Euclidean	2	0.596	2.3152	3.6315	0.1782
	Manhattan	2	0.8487	3.4028	5.6587	0.1014
200	Euclidean	2	0.6503	2.5127	3.9595	0.176
	Manhattan	2	0.5639	2.1428	3.341	0.1911
500	Euclidean	2	0.467	1.7195	2.5915	0.2354
	Manhattan	2	0.5771	2.3445	3.7146	0.165
50	Euclidean	3	0.8668	3.6491	5.9238	0.0989
	Manhattan	3	0.6097	2.3471	3.7466	0.1729
100	Euclidean	3	0.5747	2.2795	3.5218	0.1665
	Manhattan	3	0.611	2.3693	3.7529	0.1869
200	Euclidean	3	0.6651	2.6675	4.3162	0.1789
	Manhattan	3	0.5173	1.9558	3.0337	0.2456
500	Euclidean	3	0.5659	2.1447	3.365	0.2008
	Manhattan	3	0.5821	2.2053	3.4761	0.1804
	基準模型		0.339	1.9591	3.4041	0.061
	0050 ETF		0.0725	0.3993	0.5865	0.1236
	台灣加權報酬指數		0.1433	0.8258	1.1821	0.1335

DOI:10.6814/NCCU202200715

如圖 10、圖 11、圖 12 及圖 13 所示，可以看到不同的分群數量隨著時間的推移，在 Fuzzy c-mdeoids 聚類模型也有越來越顯著的差異，不過隨著距離度量及模糊性參數的調整，並沒有一個分群數量顯著優於他者；距離度量的選擇同樣對於績效也有影響，但隨著分群數量及模糊性參數的調整，我們也無法找到一個距離度量絕對優於另一個；模糊性參數也同樣如此。我們推測使用 Fuzzy c-mdeoids 聚類模型是具有一定的效果，但參數之間具有較強的交互作用，例如：將模糊性參數調高，樣本點與中心點的距離將會加大影響中心點的設置，其中距離的計算又與距離度量有關，在這之中，分群的數量也會影響群集內雜訊的含量，造成參數之間不易得到最佳解。

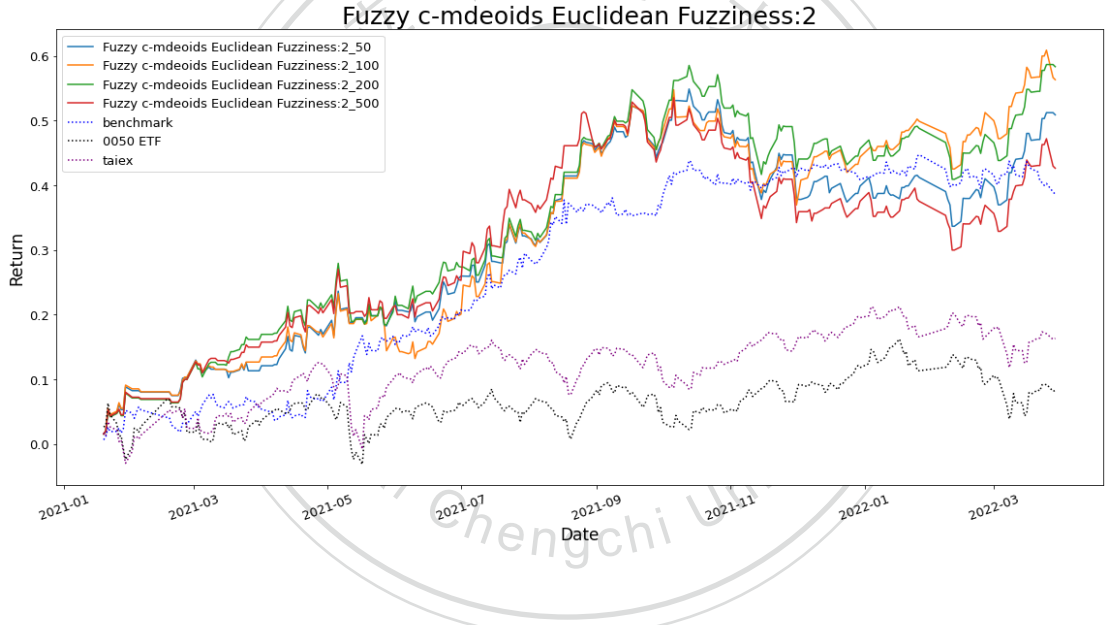

圖 10 使用歐式距離及模糊性參數 2 與不同群集數量之權益曲線

DOI:10.6814/NCCU202200715

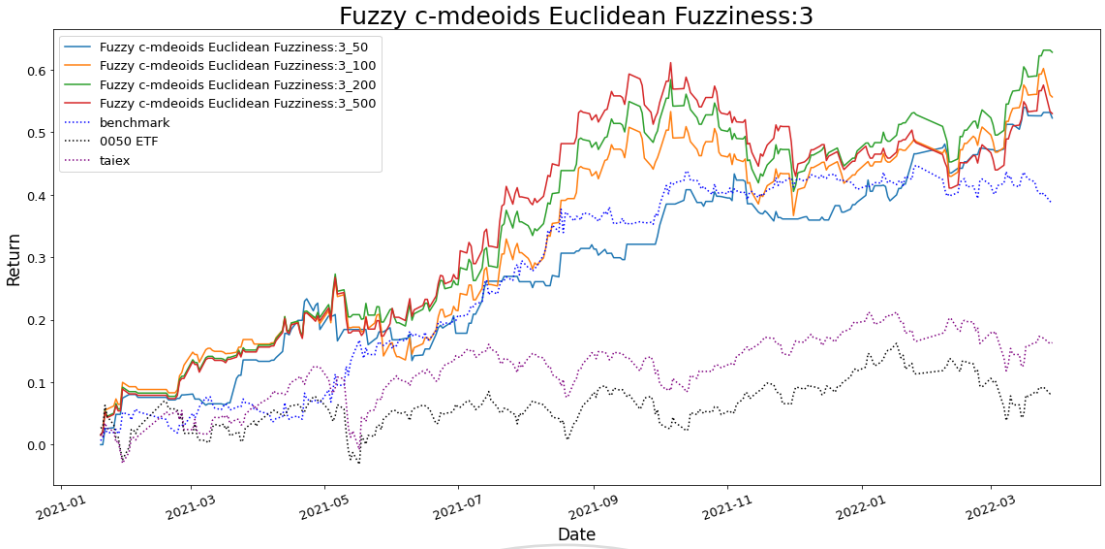

圖 11 使用歐式距離及模糊性參數 3 與不同群集數量之權益曲線

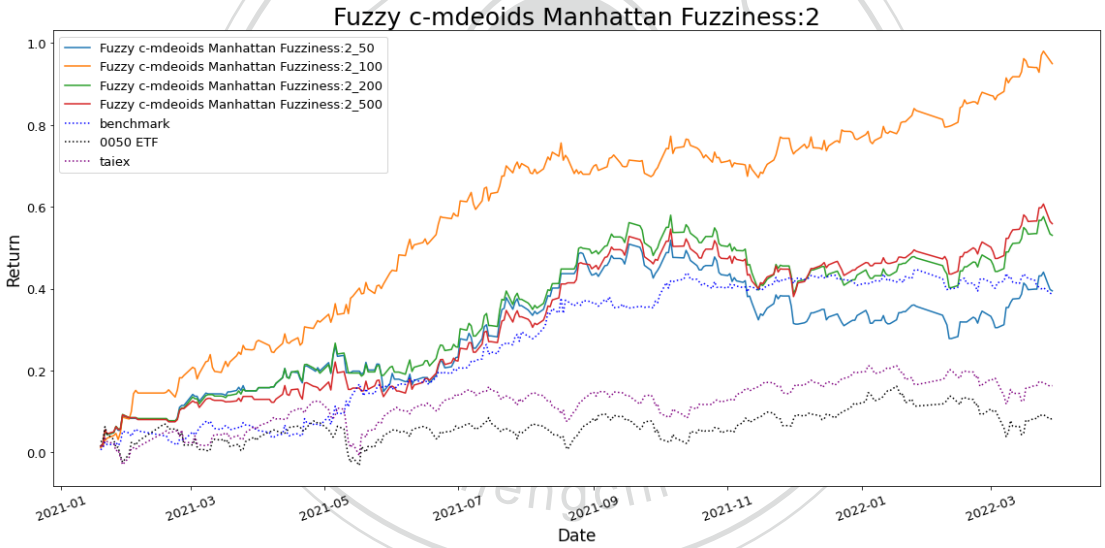

圖 12 使用曼哈頓距離及模糊性參數 2 與不同群集數量之權益曲線

DOI:10.6814/NCCU202200715

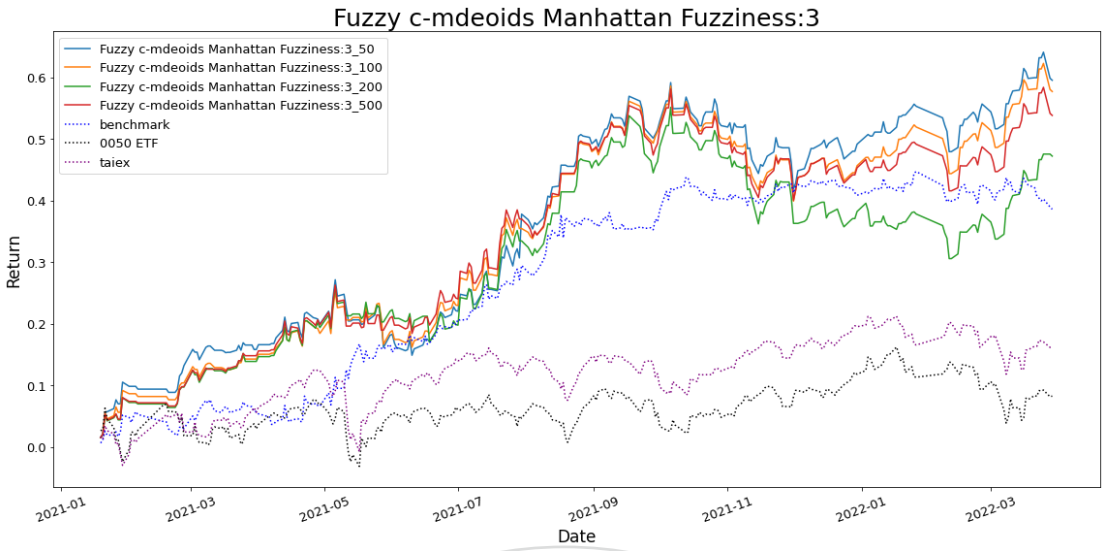

圖 13 使用曼哈頓距離及模糊性參數 3 與不同群集數量之權益曲線

三、模型間比較

這個部份我們比較兩個模型對於績效的差異，為了簡化，在 K-mdeoids 聚類模型選取兩組表現較好的參數結果，Fuzzy c-mdeoids 聚類模型同樣選取兩組參數結果進行比較，由於距離度量對於各個模型影響不一，我們統一使用歐式距離進行比較，如表 3 所示。

DOI:10.6814/NCCU202200715

表 3 模型間績效比較

聚類模型	集群數量	距離度量	模糊性參數	年化報酬率	夏普比率	索提諾比率	最大回撤
K-mdeoids	100	Euclidean	無	0.704	3.5228	6.1245	0.0475
K-mdeoids	200	Euclidean	無	0.7961	4.0512	7.5423	0.0431
Fuzzy c-mdeoids	100	Euclidean	2	0.596	2.3152	3.6315	0.1782
Fuzzy c-mdeoids	100	Euclidean	3	0.5747	2.2795	3.5218	0.1665
Fuzzy c-mdeoids	200	Euclidean	2	0.6503	2.5127	3.9595	0.176
Fuzzy c-mdeoids	200	Euclidean	3	0.6651	2.6675	4.3162	0.1789
基準模型				0.339	1.9591	3.4041	0.061
0050 ETF				0.0725	0.3993	0.5865	0.1236
台灣加權報酬指數				0.1433	0.8258	1.1821	0.1335

DOI:10.6814/NCCU202200715

其中，K-mdeoids 聚類模型在績效表現上大部分是優於 Fuzzy c-mdeoids 聚類模型，無論是夏普比率、索提諾比率、最大回徹等皆是如此；同樣地，兩個模型的表現皆可以勝過台灣 50 ETF 與台灣加權報酬指數。我們認為使用兩個模型對於績效提升皆有一定程度的幫助，會造成其中的差異是因為兩者邏輯上最大的差異——是否保留距離的資訊，K-mdeoids 聚類模型判斷樣本點的集群類別是絕對的；反之 Fuzzy c-mdeoids 聚類模型則是以機率的方式呈現。由於金融數據本身就存在大量的雜訊，加上本研究使用 1 分鐘 k 棒的報酬時間序列，在短時間的報酬波動更是如此，保留較多資訊的 Fuzzy c-mdeoids 聚類模型因大量雜訊，造成分群結果反而差強人意；反觀 K-mdeoids 聚類模型去除許多不必要的資訊，使得能有較好的分群結果。我們認為如果資料本身存在較多雜訊時，我們應該選擇 K-mdeoids 聚類模型作為研究的方向；倘若能將資料進行更進一步的前處理降低雜訊，或選擇雜訊較少的資料作為輸入值，我們再將 Fuzzy c-mdeoids 聚類模型納入模型選擇的範圍。

DOI:10.6814/NCCU202200715

第五章 結論與展望

第一節 研究結論

本文研究以 2020 年 4 月至 2022 年 3 月台股所有上市櫃股票 1 分鐘的 k 的報酬時間序列，藉由動態時間校正計算各個報酬時間序列的距離，接著用聚類分析將相似的時間序列分群，並嘗試建構當日沖銷的放空交易策略。

其中聚類分析使用 K-medoids 聚類與 Fuzzy c-medoids 聚類進行比較，分群數量嘗試 50,100,200,500 個群集，距離度量嘗試歐式距離與曼哈頓距離，Fuzzy c-medoids 聚類中的模糊性參數嘗試 2,3。實證發現，大部分的參數組合，使用動態時間校正搭配聚類分析能有效選出有異常報酬的股票，無論是累積報酬率、夏普比率、索提諾比率等，其績效表現皆比基準模型、台灣加權報酬指數、台灣元大 50ETF 表現更好，說明股票的走勢分群，並挑選適當的群集可以有效提高報酬並降低風險。

除此之外，本文比較不同的分群方法及不同的參數輸入發現，K-medoids 聚類又以 100 與 200 個集群表現較好，推論 50 及 500 個群集分別有分類不足，群集內雜訊內增加以及分類過多，產生許多無效群集等問題；Fuzzy c-medoids 聚類則是參數交互作用較強，沒能找到較突出的參數組合，但同樣具有勝過基準模型的能力；比較兩個模型，K-medoids 聚類表現普遍優於 Fuzzy c-medoids 聚類，推測儘管相對於 K-medoids 聚類，Fuzzy c-medoids 聚類可以保留樣本間的距離等更多資訊，但用於短期的金融資料同時也包含許多雜訊，造成分群結果不盡理想。

DOI:10.6814/NCCU202200715

第二節 未來展望

本文輸入的資料僅有 1 分鐘 k 報酬時間序列，由於頻率較短本身就含有大量的雜訊，因此未來可以考慮使用較長頻率的報酬時間序列，如此不僅能解決雜訊過多的問題外，還可以減少頻繁進出場的摩擦成本。除此之外，僅用價格資訊預測未來預期報酬，可能存在輸入資訊量不足的問題，造成預測效果不良，加上更多股票相關的資訊可能會有更好的預測表現，例如：公司財務指標、籌碼資料等。

本文使用動態時間校正搭配 Sakoe & Chiba (1978)提出的約束方法，對於模型訓練效率有所提升，但實證結果仍需花費超過半小時才能訓練一組模型，造成超參數調整仍舊不容易，未來可以加上由 Kim et al. (2001)所提出的的 LB_Kim、Yi et al. (1998)提出的 LB_Yi 或 Keogh & Ratanamahatana (2005)提出的 LB_Keogh 作為模型訓練的限制，提升模型訓練的效率。

DOI:10.6814/NCCU202200715

參考文獻

1. Berndt, D. J., & Clifford, J. (1994). *Using dynamic time warping to find patterns in time series.* Paper presented at the KDD workshop.

2. Cuong, N. A., Mai, D. S., Hop, D. T., Ngo, L. T., & Long, P. T. (2021). *Fuzzy C-Medoids Clustering Based on Interval Type-2 Inituitionistic Fuzzy Sets.* Paper presented at the 2021 RIVF International Conference on Computing and Communication Technologies (RIVF).

3. Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on acoustics, speech, and signal processing, 23*(1), 67-72.

4. Izakian, H., Pedrycz, W., & Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence, 39*, 235-244.

5. Jin, X., & Han, J. (2010). K-Medoids Clustering, Encyclopedia of Machine Learning. In: Springer US.

6. Kaufman, L., & Rousseeuw, P. J. (1990). Partitioning around medoids (program pam). *Finding groups in data: an introduction to cluster analysis, 344*, 68-125.

7. Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and information systems, 7*(3), 358-386.

8. Kim, S.-W., Park, S., & Chu, W. W. (2001). *An index-based approach for similarity search supporting time warping in large sequence databases.* Paper presented at the Proceedings 17th international conference on data engineering.

9. Krishnapuram, R., Joshi, A., Nasraoui, O., & Yi, L. (2001). Low-complexity fuzzy relational clustering algorithms for web mining. *IEEE transactions on Fuzzy Systems, 9*(4), 595-607.

10. Labroche, N. (2010). *New incremental fuzzy c medoids clustering algorithms.* Paper presented at the 2010 Annual Meeting of the North American Fuzzy

DOI:10.6814/NCCU202200715

Information Processing Society.

11. Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition, 44*(3), 678-693.

12. Ratanamahatana, C. A., & Keogh, E. (2005). *Three myths about dynamic time warping data mining.* Paper presented at the Proceedings of the 2005 SIAM international conference on data mining.

13. Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on acoustics, speech, and signal processing, 26*(1), 43-49.

14. Sammut, C., & Webb, G. I. (2017). *Encyclopedia of machine learning and data mining*: Springer Publishing Company, Incorporated.

15. Torra, V. (2015). *On the selection of m for Fuzzy c-Means.* Paper presented at the IFSA-EUSFLAT.

16. Yi, B.-K., Jagadish, H. V., & Faloutsos, C. (1998). *Efficient retrieval of similar time sequences under time warping.* Paper presented at the Proceedings 14th International Conference on Data Engineering.

DOI:10.6814/NCCU202200715