

國立交通大學
電機與控制工程研究所
碩士論文

應用時間序列相似度量測方法
於異常偵測與分類

Applying Time-Series Similarity Measurement
to Fault Detection and Classification

研究生：洪達文

指導教授：周志成 博士

中華民國九十五年九月

應用時間序列相似度量測方法
於異常偵測與分類

Applying Time-Series Similarity Measurement
to Fault Detection and Classification

研究生：洪達文

Student：Ta-Wen Hung

指導教授：周志成

Advisor：Dr. Chi-Cheng Jou



A Thesis

Submitted to Department of Electrical and Control Engineering
College of Electrical Engineering
National Chiao-Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Electrical and Control Engineering
September 2006
Hsinchu, Taiwan, Republic of China

中華民國九十五年九月

應用時間序列相似度量測方法 於異常偵測與分類

學 生：洪達文

指導教授：周志成 博士

國立交通大學電機與控制工程研究所

摘 要

本論文主要探討如何在滿足正常狀態之工作範圍的條件下，偵測出異常的時間序列資料並將其分類，且該方法與系統對時間序列位移量與雜訊大小的敏感性越小越好。相較於傳統的工作範圍偵錯法只能偵測出超出工作範圍的異常，以及主成份分析法容易損失資料資訊，造成微小異常的不易偵測與分類，本論文以直接量測時間序列的相似度來改進這些缺失。本研究以半導體製程中，質流控制器的五種異常狀態為例，希望在滿足正常工作範圍的條件下，能將異常訊號偵測出來，並可將正常與異常共六種類別分類。首先，使用歐基里德距離量測法與動態時間扭曲法，測量出時間序列間的相似度。接著，利用多重維度量測法的二維知覺圖與 F -試驗值初步評估系統成效。之後，再使用 k -最鄰近點分類法來進一步驗證偵錯與分類結果，並比較不同相似度量測方法的系統成效差異。同樣的系統除了可應用在本論文所舉的半導體製程例子之外，亦可應用於與時間序列有關的議題上，例如金融與醫學等領域。

關鍵字：時間序列、工作範圍、異常偵測與分類、相似度、歐基里德距離量測法、動態時間扭曲法、多重維度量測法、知覺圖、 F -試驗值、 k -最鄰近點法。

Applying Time-Series Similarity Measurement to Fault Detection and Classification

Student : Ta-Wen Hung

Advisor : Dr. Chi-Cheng Jou

Department of Electrical and Control Engineering
National Chiao Tung University

Abstract

The purpose of this thesis is to detect and classify abnormal time-series data under normal working range. In addition, the less sensitive the offset and noise of time-series data is, the better the system is. The traditional working-range-detection method is just used to detect faults over normal working range, and using principal components analysis is not easy to detect and classify little abnormality because it loses information of data. The thesis improves these shortcomings by measuring the similarity of time-series data directly. This research takes five faults of mass flow controller in semiconductor manufacturing process for example. It is expected that faults can be detected and classified under normal working range. First, we measure similarity between different time-series data with Euclidean distance metric and dynamic time warping. Second, we make use of two-dimensional perceptual map of multidimensional scaling and F-test to evaluate the preliminary performance of the system. And then, we validate the results of fault detection and classification by using k -nearest neighbors classification, and comparing the performance of the system using different similarity measurement. The same system can be applied not only to semiconductor manufacturing process in this thesis, but to the issues about time-series data in financial and medical domain.

Keywords : Time-series, Working Range, Fault Detection and Classification, Similarity, Euclidean Distance Metric, Dynamic Time Warping, Multidimensional Scaling, Perceptual Map, F -test, k -Nearest Neighbors.

誌 謝

由衷感謝指導教授周志成老師的細心指導與教誨，除了學術研究上的訓練與知識外，人生價值觀的建立以及對事物的觀察與見解等，更是對我有著莫大幫助。同時亦感謝楊谷洋老師與陳永平老師在口試時給予的指導與建議，使本論文研究能更形完善。感謝實驗室的志強與似修學長，建威與丞師，以及星戎和任宏學弟，他們在課業與生活上的支持與協助，使我在碩士生涯能有著充實經歷與回憶。最後，更要感謝家人與親朋好友的支持與關懷，能和我一起分享喜怒哀樂，一路走來縱使有小顛簸，亦能順順利利的度過這一階段。僅將此論文獻給所有我所認識的人，謝謝你們！



目 錄

中文摘要	i
英文摘要	ii
誌謝	iii
目錄	iv
表目錄	vi
圖目錄	vii
1 序論	1
1.1 研究動機與目的	1
1.2 問題特徵與文獻方法	2
1.3 研究方法與系統架構	6
1.4 論文架構	8
2 統計方法與模型	9
2.1 主成份分析法	9
2.1.1 特性簡介	9
2.1.2 運算機制	11
2.2 多重維度量法	14
2.2.1 知覺圖	15
2.2.2 運算機制	17



2.3	k -最鄰近點分類器	21
2.3.1	未加權式 k -最鄰近點	21
2.3.2	加權式 k -最鄰近點	22
2.4	定義評估函數	24
3	時間序列之相似度量測	27
3.1	歐基里德距離量測法	27
3.2	動態時間扭曲法	28
3.2.1	特性簡介	28
3.2.2	視窗式動態時間扭曲法	30
3.2.3	最小累進距離式動態時間扭曲法	32
4	實驗與討論	33
4.1	資料產生與實驗流程	33
4.1.1	資料產生	33
4.1.2	實驗流程	38
4.2	類別內實驗	40
4.3	類別間實驗	47
4.4	驗證分類結果	51
5	研究結論	58
	參考文獻	61



表 目 錄

2.1	多重維度量法範例	18
2.2	不同訓練資料樣本數的分類成效	24
4.1	N2與不同位移量下的各系統 F -試驗值	43
4.2	N8與不同位移量下的各系統 F -試驗值	46
4.3	不同雜訊大小下的各系統 F -試驗值	50
4.4	N2之各系統分類錯誤率	51
4.5	N8之各系統分類錯誤率	52
4.6	N1k3之各系統分類錯誤統計	54
4.7	N5k3之各系統分類錯誤統計	54
4.8	N9k3之各系統分類錯誤統計	55
4.9	EDM系統之分類錯誤統計	55
4.10	DTW系統之分類錯誤統計	55
4.11	N1k3-EDM之詳細分類表	56
4.12	N1k3-DTW之詳細分類表	56
4.13	N9k3-EDM之詳細分類表	57
4.14	N9k3-DTW之詳細分類表	57

圖目錄

1.1	正常工作範圍範例	2
1.2	MFC異常範例	3
1.3	時間序列維度縮減範例	4
1.4	PCA 範例	5
1.5	系統架構流程圖	6
1.6	各種分類器的分類錯誤率	7
2.1	主成份分析法流程圖	9
2.2	散佈圖範例	10
2.3	第一主成份範例	11
2.4	知覺圖範例	15
2.5	知覺圖的分類	16
2.6	多重維度度量法流程圖	17
2.7	多重維度度量法範例	18
2.8	距離與內積之間的幾何關係	18
2.9	相似性與相異性矩陣的分類	19
2.10	k NN 分類器	22
2.11	DW- k NN 分類器	23
2.12	k 值選擇	23
2.13	分類範例	25
3.1	歐基里德距離量測法範例	27

3.2	歐基里德距離量測法相似度分類範例	28
3.3	相似度運算圖解	29
3.4	動態時間扭曲法相似度分類範例	29
3.5	動態時間扭曲法路徑範例	30
4.1	資料產生流程圖	34
4.2	正常資料時間序列之產生	35
4.3	正常狀態之工作範圍	35
4.4	異常類別資料樣板	36
4.5	異常類別資料之時間序列	37
4.6	整體實驗流程圖	39
4.7	類別內實驗流程圖	40
4.8	類別內實驗之控制變數	41
4.9	類別內實驗之小雜訊時間序列	41
4.10	N2S4之知覺圖	42
4.11	N2與不同位移量下的各系統成效	44
4.12	類別內實驗之大雜訊時間序列	44
4.13	N8S6之知覺圖	45
4.14	N8與不同位移量下的各系統成效	46
4.15	類別間實驗流程圖	47
4.16	類別間實驗之異常時間序列	48
4.17	N5之知覺圖	49
4.18	不同雜訊大小下的各系統成效	50
4.19	驗證流程圖	51
4.20	N2之各系統分類成效	52
4.21	N8之各系統分類成效	53
4.22	驗證之控制變數	53

第 1 章

序論

1.1 研究動機與目的

在資料探勘 (data mining) 中, 尋找出異常 (abnormal) 的資料是個重要的議題, 尤其能否偵測出發生在定義的正常工作範圍 (normal working range) 內之微小異常, 更是導致資料探勘系統成效好壞的重要因素之一。而在偵測出異常訊號之後, 我們更進一步地想知道如何能將該異常分類, 以期能針對該異常類別累積經驗, 如此當該異常再次發生時即可有效地處理。在業界中, 諸如半導體工業、金融業、醫學等, 其領域上的資訊常會以時間序列 (time-series) 的資料方式呈現, 亦即“時間”是其資料的一個重要變數。所謂時間序列, 意指每單位時間所測量或得到的數據, 將其按時間的先後發生順序排列而成的資料即稱之。由時間序列的特性可知, 單一時間點上的資料有時並不是資料探勘的焦點所在, 反而是某段時間內的資料趨勢 (trend) 更引人關注, 因此尋找一個能有效判別不同時間序列間之趨勢差異的量測方法, 成爲本研究的主要方向。

在半導體工業中, 各種製程都有與時間相關的參數存在, 例如氣體濃度對時間的變化等等。製程變異對產品品質或良率有很重要的影響, 如能正確的偵測出異常並將之分類, 除了能改善產品品質與良率外, 亦可使製造成本降低、提高生產率並進而獲取更多利潤[1]。異常偵測與分類 (Fault Detection and Classification, FDC) 能幫助工程師更好理解製程過程中或設備的細微差異, 進而快速且正確的解決問題所在, 降低異常所造成的損失。

同樣的, 在金融業中, 經濟、股市等走勢亦是大眾平日關注的議題。以經濟走勢為例, 如能及早偵測出走勢異常, 則可預警 (early warning) 金融危機的發生, 而若能將異常分類, 則可就經驗法而對症下藥, 進而減低甚至避免金融危機所帶來的影響。再以股市走勢為例, 投資人無不希望有一能預測未來股價波動的方法供其作投資參考, 若將股價下跌視為一種異常來處理, 則發展一套偵測股市走勢異常的系統就有其可行與應用性[2]。而在醫學上, 定時為病患量測相關醫學數據是醫療的方式之一, 這些數據即為一種時間序列。以前列腺癌 (prostate cancer) 為例, 臨床資料顯示, 與此癌症有關的診斷包括囊狀滲透率 (cystic osmosis)、前列腺增值量 (prostatic hyperplasia) 與葛利森值 (Gleason grading) 等, 如能偵測這些時間序列的異常並加以分類, 則醫師的診斷與治療將更有效率, 進而降低癌症的發生機率[3]。

綜觀上述, 可將問題整理如下: 發展一套系統, 使得當時間序列資料在所定義的正常工作範圍內, 有異常情況發生時 (由於在正常工作範圍內, 因此該異常時間序列很可能與正常情況時的時間序列相似), 仍能偵測出該異常並同時將其分類, 且希望此系統對時間序列的位移量 (offset) 及雜訊 (noise) 大小的敏感性 (sensitivity) 越小越好, 如此系統才能更為耐用 (robust)。

1.2 問題特徵與文獻方法

舊有的偵錯方法之一 (在本論文中稱之為“工作範圍偵錯法”), 是定義出正常狀態的工作範圍, 如圖 1.1 所示, 至於該如何定義工作範圍則有多種方式, 通常是以資料

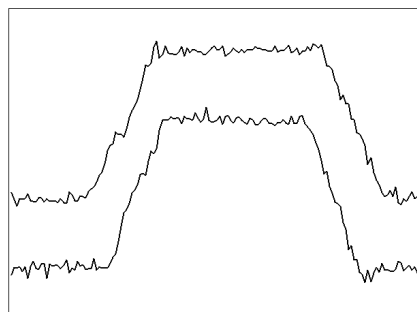


圖 1.1: 正常工作範圍範例

集的均值加(減)上一值,例如3個標準差。當訊號超出該範圍時,警報系統(warning system)就發出異常警告;換句話說,若時間序列未超出該工作範圍,則不管其是否為異常情況,警報系統都不會有所反應[4]。本論文研究的資料型態為單變量的時間序列,問題著重在正常工作範圍下的異常偵測,因此前述舊有方法明顯不適用。再者,除了偵錯之外,如有多種異常情形,尚需對異常加以分類,因此必須尋求另一更好的方法。

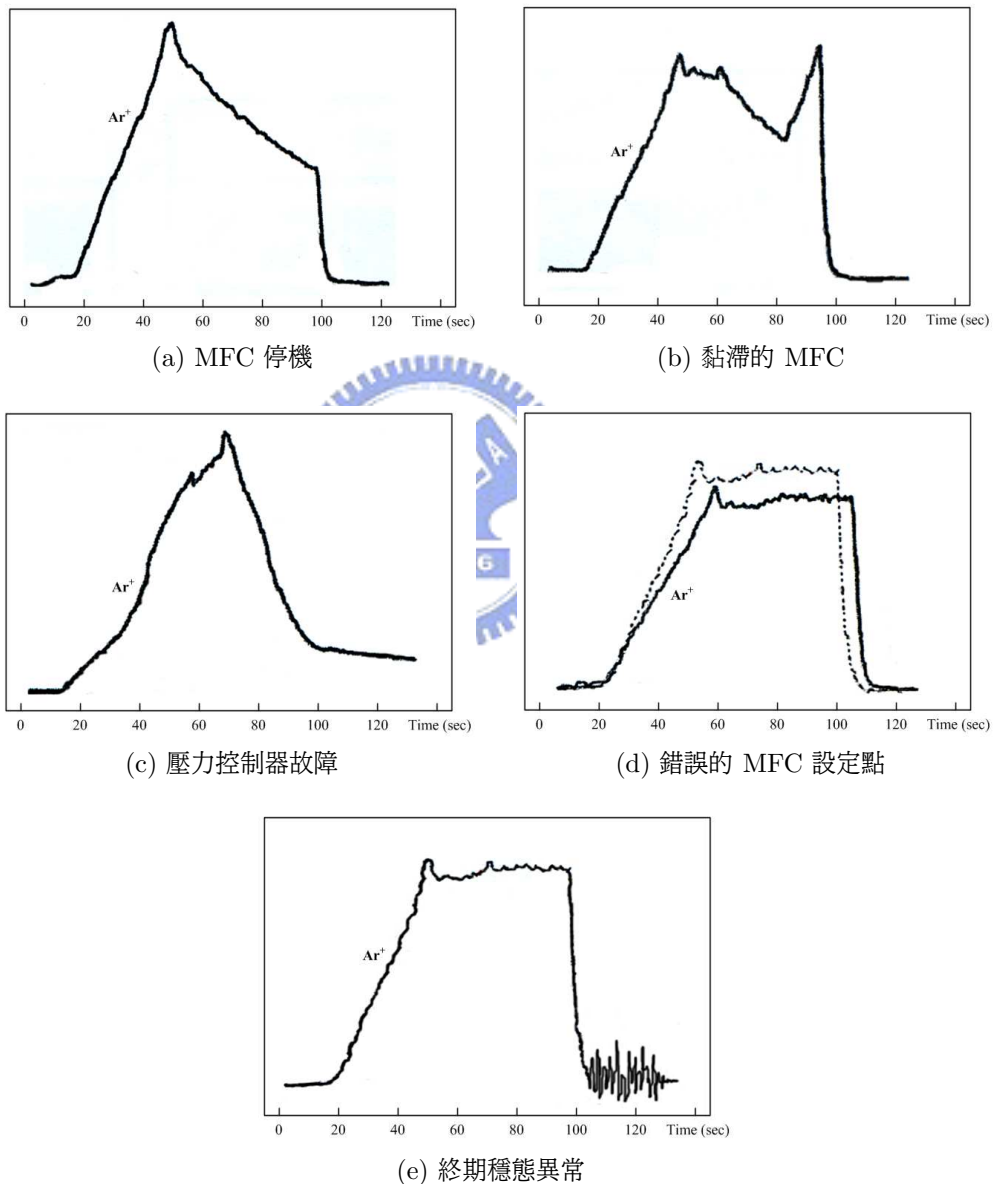


圖 1.2: MFC 異常範例

本論文以半導體工業的製程資料為例, 在晶片的製造過程中需使用到多種製程技術, 例如微影 (photolithography)、蝕刻 (etching)、化學氣相沉積 (chemical vapor deposition)、物理氣相沉積 (physical vapor deposition)、擴散(diffusion)、離子植入 (ion implantation) 與氧化 (oxidation) 等。其中, 除了黃光微影製程之外, 幾乎所有的單位製程, 都是藉由多種製程氣體的化學或物理反應來達成, 並且透過質流控制器 (Mass Flow Controller, MFC) 對各種製程氣體進入反應室 (chamber) 的質量加以量測與調整, 進而控制反應的進行。由於製程的進行與良率取決於 MFC 是否能依照設定值, 穩定且精確地控制反應氣體從管線流進反應室內的質量, 所以在半導體製程中, MFC 扮演了相當關鍵性的角色[5, 6]。圖 1.2 展示了 MFC 可能發生的各種異常情況, 例如 MFC 停機 (MFC shutdown)、黏滯的 MFC (sticky MFC)、壓力控制器故障 (pressure controller failure)、錯誤的 MFC 設定點 (wrong MFC set point)、終期穩態異常 (abnormal final steady-state) 等, 本研究將以此圖作為虛擬資料建構的參考[7, 36]。圖 1.2 是以氬氣 (Ar) 為例, 橫軸為時間, 縱軸為氣體濃度, 其中圖 (d) 內的虛線代表典型的正常狀態。

從一些文獻與業界現行的使用方法可得知, 既有的系統是以主成份分析法 (Principal Components Analysis, PCA, 詳見節 2.1) 為主體。首先先對時間序列作維度縮減, 以圖 1.3 為例, 圖中長度為 130 的時間序列每 10 單位被等分成一段, 每一區段取一值作代表, 通常是以該段的各資料點均值作為取樣值, 如此原本 130 點的資料被縮減為 13 點, 可將之視為一 13 維空間中的座標值。接下來, 取出幾個變

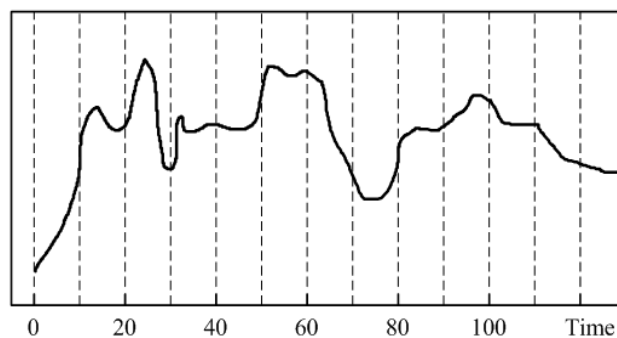


圖 1.3: 時間序列維度縮減範例

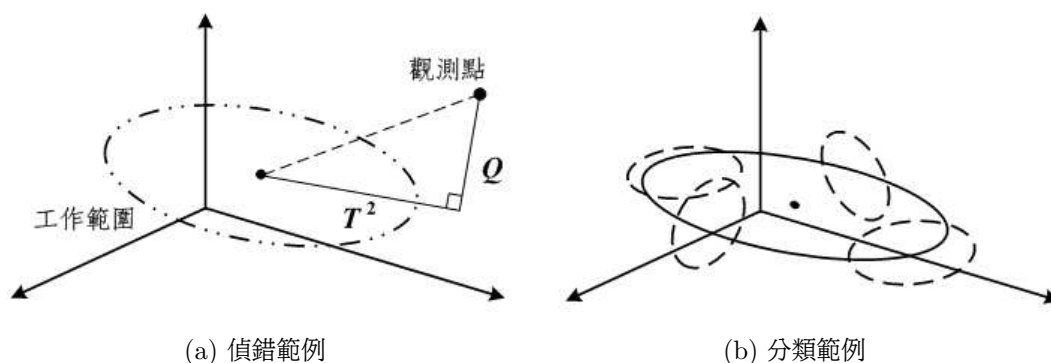


圖 1.4: PCA 範例

異量最大的變量作為主成份 (空間中的主軸), 並將時間序列轉換為維度縮減空間中的一個點, 當所有正常時間序列轉換完成後, 即可在維度縮減空間中得出一正常狀態之工作範圍。如圖 1.4 (a) 所示, 正常之工作範圍以一空間平面上的橢圓為例, 當一新的觀測點 (observation) 加入時, 可從兩方面來判斷其是否為異常資料: 第一, 觀測點投影至該平面後, 若落在工作範圍 (橢圓) 之外 (亦即 T^2 過大), 則判定為異常; 第二, 若觀測點與該平面的距離 Q 過大, 則代表觀測點處於差異過大的另一平面之上, 亦即整個資料結構產生改變, 因此判定該資料為異常。除了上述兩種情況之外, 此法皆判定新加入之時間序列為正常狀態[8, 9, 10, 11]。

但從圖 1.2 中可看出異常狀態有時非常近似於正常狀態, 以 (e) 圖為例, 整個異常時間序列只有在訊號的最末端產生變動, 若以圖 1.3 的維度縮減範例來說明, 則可能只是某一區段間有所異常, 當取了該區段的均值作為取樣值之後, 則此異常訊號經維度縮減後與正常訊號間的差異將變的非常微小, 也就不容易被偵測出來。再以圖 1.4 (b) 做說明, 圖中以 PCA 的前兩個主成份在空間中得出一平面後, 該面上的實線橢圓代表正常時間序列所建構出的正常狀態之工作範圍, 而四個虛線橢圓則代表了實際上為四種不同類別的異常訊號之範圍, 因此當異常範圍與正常範圍有交集時, 以舊有的圖 1.4 (a) 偵錯方法將無法完全偵測出來, 交集越大則偵測錯誤率越高。由於維度縮減易造成原始資料資訊的損失, 以及 PCA 偵測異常與分類的不完善, 使得文獻與業界現行的系統還能再加以改進, 因此本研究選擇直接對原

始時間序列做相似度量測，並以各個既有的異常典型訊號作為分類依據，如此即可避免資料資訊的遺失，使得微小的差異亦能予以區別出來。

1.3 研究方法與系統架構

不同於 PCA 先將觀測資料映射至所創造的維度縮減空間中再行計算，本系統直接對時間序列做相似度量測，如此即可避免使用 PCA 所造成的資料資訊遺失問題。尤其當僅有的資訊為時間序列兩兩之間的相關性 (correlation) 或相似 (similarity)、相異性 (dissimilarity) 時，本系統更較 PCA 來的合適。另外從圖 1.3 中可知，舊系統資料集內的時間序列需為相同的時間長度，因其需映射至相同的維度，但本系統在量測時間序列相似度時所使用的動態時間扭曲法 (Dynamic Time Warp-

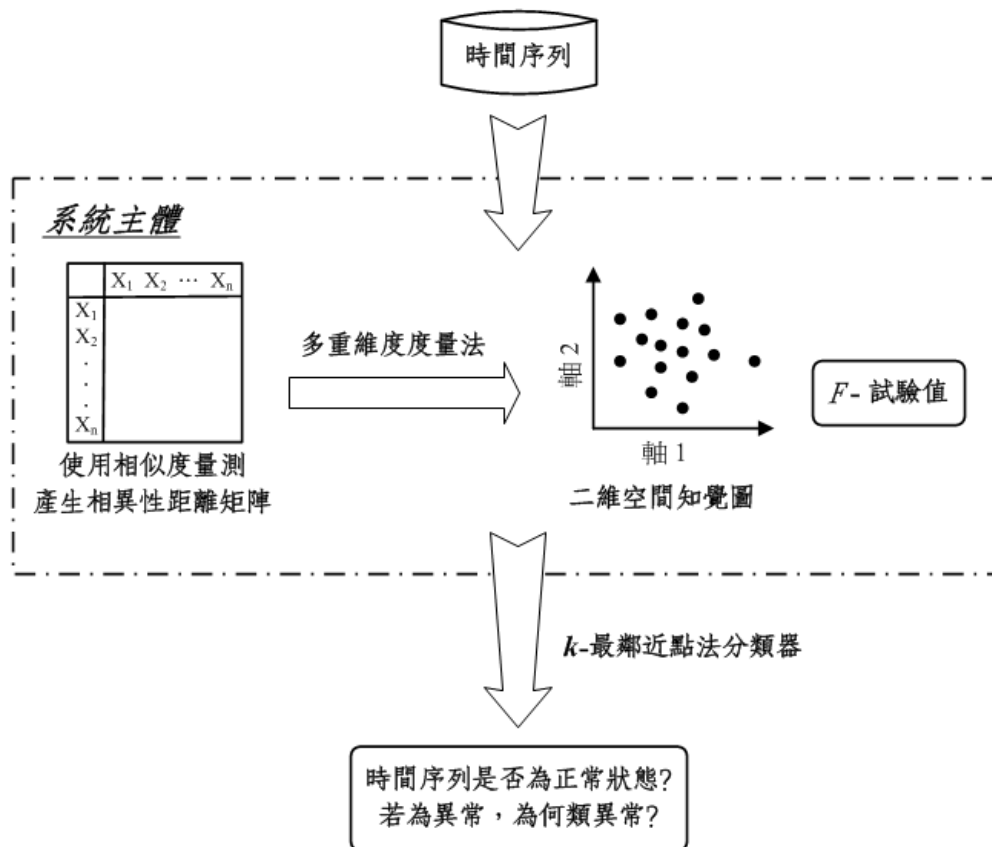


圖 1.5: 系統架構流程圖

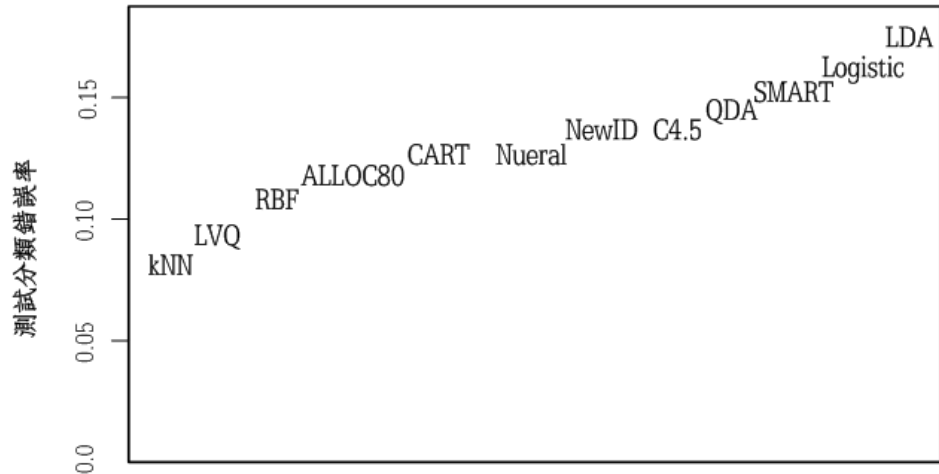


圖 1.6: 各種分類器的分類錯誤率

ing, DTW, 詳見節 3.2) 將可避免此一限制。

圖 1.5 為本研究所發展的系統架構，系統輸入資料為時間序列，輸出為多重維度度量法 (Multidimensional Scaling Method, MDS, 詳見節 2.2) 所產生的二維空間知覺圖 (perceptual map)，以及自行定義的評估函數 F -試驗值 (F -test)，再經由加權式 k -最鄰近點 (k -Nearest Neighbors, k NN) 分類器即可偵測異常與分類。

時間序列進入本系統之後，經由歐基里德距離量測法 (Euclidean Distance Metric, EDM, 詳見節 3.1)¹ 或動態時間扭曲法等相似度測量法，得出一相異性距離矩陣，此矩陣即為本系統最重要的概念，接下來是將該矩陣轉換為使用者所能理解的資料與方式 — 二維空間知覺圖與 F -試驗值。本論文主要是研究各時間序列間的相似度評比，屬於認知研究，將心理上的抽象知覺資料以數據方式，轉換成知覺圖中的座標位置，以方便分析其中的訊息，因此非常符合多重維度度量法的功能性與精神。有了知覺圖的座標值之後，便可針對觀測資料作異常偵測與分類，而除了知覺圖的直覺圖形判讀之外， F -試驗值則提供了更實際的分類成效數據。圖 1.6 是由 STATLOG 計劃 (Michie 等人, 1994)² 所實驗出的各種分類演算法之分類錯誤率

¹使用歐基里德距離量測法時，仍有與 PCA 相同的限制：不同的時間序列需有相同時間長度。

²STATLOG 是一比較各式分類演算法成效的計畫，使用超過 20 組的不同資料集來評估大約

比較, 參考該實驗結論後, 本系統選擇了 k -最鄰近點分類器³, 並加以改良為加權式 k NN [3]。

1.4 論文架構

本論文共分為五章。第二章介紹相關統計方法與模型, 除了既有文獻與業界現行的主成份分析法之外, 尚有本研究所用到的多重維度度量法與 k -最鄰近點分類器, 以及本論文所定義的評估函數 — F -試驗值。第三章介紹時間序列相似度的量測方法, 包括歐基里德距離量測法與動態時間扭曲法。第四章為實驗與討論, 包含了類別內與類別間實驗, 以及實驗結果的驗證。第五章則統整本論文研究, 並對其優缺點與未來發展做出探討。



20 種的分類演算法。

³當欲決定的分類邊界非常不規則時, k NN 仍能有效分類。

第 2 章

統計方法與模型

2.1 主成份分析法

2.1.1 特性簡介

如圖 2.1 所示，主成份分析法 (Principal Components Analysis, PCA) 是對於所有影響因素先進行因子 (factor) 分析，以抽取所有解釋變數 (explanatory variable) 之共通因子 (common factor) 作為評量基準，而因這些共通因子間皆相互獨立，故可避免在評量因子間存在之相關性時有所偏差。接下來必須在簡單性 (維度越少越容易管理) 與完整性 (維度越多將保留越大量的可用資訊) 之間做選擇，決定要保留多少的主成份做之後的分析。最後利用所得的各因子負荷量 (factor load-

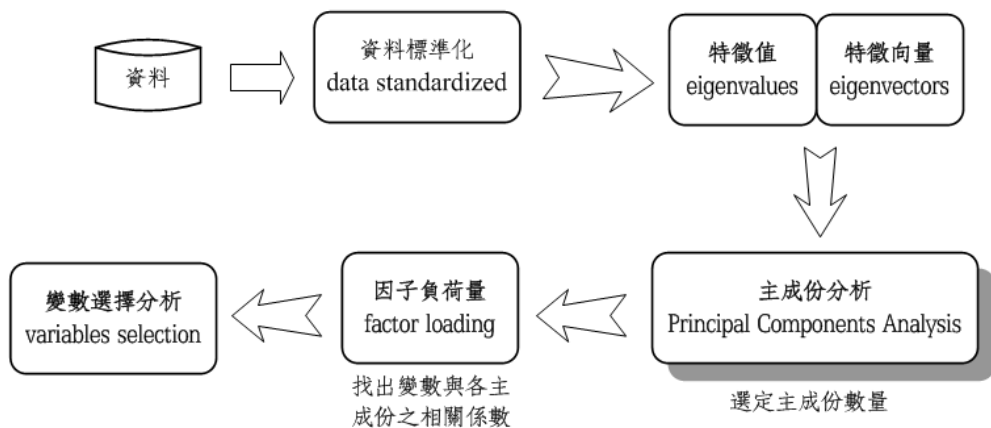


圖 2.1: 主成份分析法流程圖

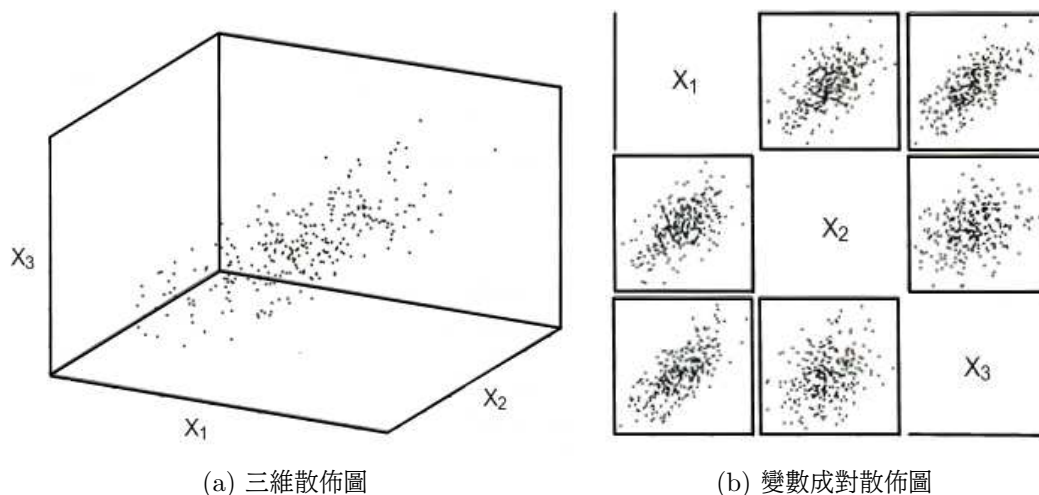


圖 2.2: 散佈圖範例

ing) 與主成份進行分析, 並描述各分類之特性[12, 13]。

PCA 主要概念是以少數主成份來表達多種變數資訊, 將多種變數予以線性組合, 使之得到的成份具有最大的變異量 (variance)¹, 所以 PCA 一般目的是對龐大資料作縮減。以 PCA 進行分析可保留原變數之分佈特性而降低變數維度, 且各成份間相互獨立可避免重覆表達原資料之特性, 以達到精簡計算分析的目的[14]。以圖 2.2 的三維散佈圖為例, 將其三個變數成對 (pairwise) 畫出散佈圖後, 可發現三變數兩兩間皆為正相關 (positive correlation), 因此我們可以說三個變數間具有共變性。而使用 PCA 將能縮減至比三個更少的維度數, 並足以解釋原有變數的資訊。常理上, 第一主成份會選擇有最大變易量的變數之線性組合², 其餘依此類推。圖 2.3 表示出第一主成份的方向, 即橢圓中的長軸。

PCA 在應用上, 具有下列優點:

- 消除變數間之共變性

變數間具有共變性, 顯示彼此間存有某種程度的重疊訊息。在迴歸 (regression) 分析中, 當自變數間具有共變性時, 以最小平方法 (least squares) 所求

¹變異量是種測量資料分散度的度量值。

²線性組合可視為多維空間中的所有點投影至單一軸上。

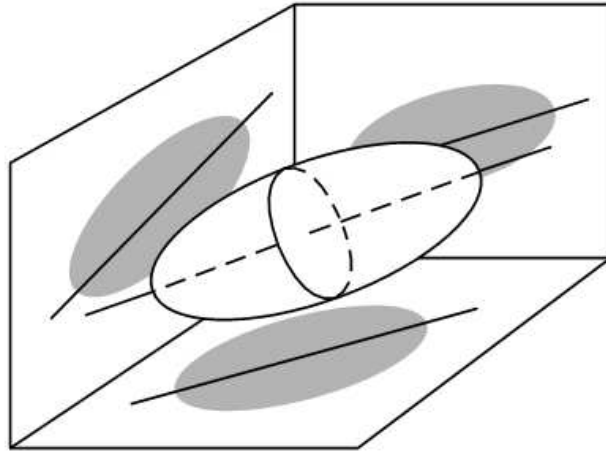


圖 2.3: 第一主成份範例

得之迴歸係數值很大, 其推估值將不易決定, 而 PCA 選取之主成份變數則可解決上述問題。

- 降低資料維度, 並抽取重要訊息

當變數維度較高時, 其資料處理分析、特性解釋與圖表之表示均較困難, PCA 可簡化此一困擾, 即降低維度求出綜合變量。此綜合變量間相互獨立, 且可反映大部分原始變數之總變異, 再以二維或三維圖形來表示較大之主成份, 有助於特性之解釋。

- 提供合適現象解釋

較重要之主成份變數, 足可解釋原始變數較大部分之總變異量, 由前幾個較重要之因子負荷量, 可觀察出原始變數與該主成份之相關程度。

2.1.2 運算機制

一般使用 PCA 的目的, 是找出原始變數 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]$ 間有著最大變異量的線性組合。假設 \mathbf{X} 已經過標準化 (standardized)³, 並以 $\mathbf{u} = (u_1, u_2, \dots, u_p)'$ 代表線性組合, 則我們的目標是選擇 \mathbf{u} 使得 $\mathbf{z} = \mathbf{X}\mathbf{u}$ 有著最大化的變異量, 可表

³標準化是將各變數資料轉換為零平均值 (zero mean) 與單位標準差 (unit standard deviation)。

示成下式

$$\text{var}(\mathbf{z}) = \frac{1}{n-1} \mathbf{u}' \mathbf{X}' \mathbf{X} \mathbf{u} \quad (2.1)$$

因爲 \mathbf{X} 是標準化的, 因此 $1/(n-1) \mathbf{X}' \mathbf{X}$ 即爲樣本相關性矩陣 \mathbf{R} , 故有

$$\text{var}(\mathbf{z}) = \mathbf{u}' \mathbf{R} \mathbf{u} \quad (2.2)$$

式 (2.2) 中, \mathbf{u} 可設爲任意大, 如此將導致 $\text{var}(\mathbf{z})$ 爲無限大。因此我們必須再加一項限制, 即 \mathbf{u} 向量具有單位長度, $\mathbf{u}' \mathbf{u} = 1$ 。

我們可利用拉格朗 (Lagrangian) 形式來解這個限制最佳化的問題[15], 令

$$L = \mathbf{u}' \mathbf{R} \mathbf{u} - \lambda (\mathbf{u}' \mathbf{u} - 1) \quad (2.3)$$

其中 λ 稱爲拉格朗乘數 (Lagrange multiplier), 當 $\mathbf{u}' \mathbf{u} = 1$ 的限制不滿足時, 可藉 λ 的選擇來彌補。將 L 對 \mathbf{u} 取導數可得

$$\frac{\partial L}{\partial \mathbf{u}} = 2 \mathbf{R} \mathbf{u} - 2 \lambda \mathbf{u} \quad (2.4)$$

將式 (2.4) 設爲零, 則

$$\mathbf{R} \mathbf{u} = \lambda \mathbf{u} \quad \text{或} \quad (\mathbf{R} - \lambda \mathbf{I}) \mathbf{u} = 0 \quad (2.5)$$

其中 \mathbf{I} 爲 $p \times p$ 的單位矩陣。式 (2.5) 爲一特徵值 (eigenvalue, λ) 與特徵向量 (eigenvector, \mathbf{u}) 的特殊架構, 當矩陣 \mathbf{R} 爲滿秩 (full rank) 時, 則運算結果將包含 p 個正特徵值 (因 \mathbf{R} 爲一正定矩陣) 與其相關聯之特徵向量。將式 (2.5) 代入

(2.2) 可得

$$\text{var}(\mathbf{z}) = \mathbf{u}'\mathbf{R}\mathbf{u} = \mathbf{u}'\lambda\mathbf{u} = \lambda\mathbf{u}'\mathbf{u} = \lambda \quad (2.6)$$

因此, λ 即為相關聯主成份的變異量。我們使用矩陣 \mathbf{D} 來代表主成份的對角化共變異 (diagonal covariance) 矩陣, $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, 設 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ 。

所有主成份的總變異量為 $\text{trace}(\mathbf{D})$, trace 是加總矩陣對角元素 (element) 的函數。因為 $\mathbf{D} = \mathbf{U}'\mathbf{R}\mathbf{U}$, 所以總變異量亦可表示為 $\text{trace}(\mathbf{U}'\mathbf{R}\mathbf{U})$, 其中 \mathbf{U}' 、 \mathbf{R} 、 \mathbf{U} 皆為方陣 (square matrix)。因為 \mathbf{U} 滿秩且矩陣內的特徵向量相互正交 (orthogonal), 又 $\mathbf{U}\mathbf{U}' = \mathbf{I}$, 故得 $\text{trace}(\mathbf{U}'\mathbf{R}\mathbf{U}) = \text{trace}(\mathbf{R}\mathbf{U}\mathbf{U}') = \text{trace}(\mathbf{R})$, 意即總變異量為相關性矩陣的對角元素和。而由前 k 項最大主成份可解釋原變數總變異量之百分比為

$$r = \frac{1}{p} \sum_{i=1}^k \lambda_i \quad (2.7)$$



另一個有用的主成份解是主成份 \mathbf{Z} 與標準化原始資料 \mathbf{X} 間的相關性矩陣, 當我們知道 \mathbf{X} 間的樣式關係時, 該相關性矩陣可幫助解釋 \mathbf{Z} 。考慮

$$\text{corr}(\mathbf{X}, \mathbf{Z}) = \frac{1}{n-1} \mathbf{X}'\mathbf{Z}_s \quad (2.8)$$

其中 $\mathbf{Z}_s = \mathbf{Z}\mathbf{D}^{-1/2}$ 是標準化後的主成份矩陣, 將 \mathbf{Z} 以 $\mathbf{X}\mathbf{U}$ 取代可得

$$\begin{aligned} \text{corr}(\mathbf{X}, \mathbf{Z}) &= \frac{1}{n-1} \mathbf{X}'\mathbf{X}\mathbf{U}\mathbf{D}^{-1/2} = \mathbf{R}\mathbf{U}\mathbf{D}^{-1/2} \\ &= (\mathbf{U}\mathbf{D}\mathbf{U}')\mathbf{U}\mathbf{D}^{-1/2} = \mathbf{U}\mathbf{D}^{1/2} \end{aligned} \quad (2.9)$$

定義因子負荷量 \mathbf{F} 為

$$\mathbf{F} = \mathbf{U}\mathbf{D}^{1/2} \quad (2.10)$$

在前 k 個主成份中, 變數 X_i 所佔之變異量表示為 $\sum_{j=1}^k f_{ij}^2$, 其中 f_{ij} 為 \mathbf{F} 中的 X_i 與 Z_j 之相關性。當 $k = p$ 時 (亦即所有主成份皆被保留), 則 $\sum_j f_{ij}^2 = 1$ 。

另由 $\mathbf{Z}_s = \mathbf{X}\mathbf{U}\mathbf{D}^{-1/2}$ 可得

$$\mathbf{X} = \mathbf{Z}_s\mathbf{D}^{1/2}\mathbf{U}' \quad (2.11)$$

資料矩陣 \mathbf{X} 被表示為三個較簡單矩陣的乘積, 此即為奇異值分解 (Singular Value Decomposition, SVD)[15], 因此 PCA 可藉由相關性矩陣 \mathbf{R} 的特徵值分析 (有時稱做頻譜分解, spectral decomposition) 或是資料矩陣 \mathbf{X} 的奇異值分解來完成。

對於主成份個數之決定, 目前尚無一特定標準, 部分經驗說明以 70% – 90% 之變異解釋能力作為選定之標準; 且當變數增加或觀測數量增加時, 則累積比率可取小一點。根據凱薩規則 (Kaiser's rule) 或 Anderson 的拇指定則 (thumb rule), 建議主成份的數量選取可取特徵值 (變異數) $\lambda \geq 1$ 的個數[16, 17]。另外 Cattell 提出的陡坡測試法 (scree test) 亦可作為主成份個數之選擇標準[18]。

2.2 多重維度度量法

度量化方法是指用數字表示某特性所使用的方法, 其中又可分成單維度度量法 (Unidimensional Scaling Method, UDS)⁴ 與多重維度度量法 (Multidimensional Scaling Method, MDS)[19, 20]。

⁴所謂的單維度度量法是指可用單一數值來表達某一概念, 如身高可用公分、體重可用公斤來測量, 其優點為較易於理解。常見的包括 Thurstone Scales、Likert Scales 與 Guttman Scales 等。

2.2.1 知覺圖

許多多變量分析 (multivariate analysis) 的方法可用來找出隱藏在資料背後的結構, 將研究樣本之間的關係以空間形式來描述, 也就是“知覺圖”(perceptual map)。圖 2.4 為一知覺圖範例, 圖中將原始的四類資料做過運算處理後, 投射到二維空間的知覺圖上, 如此即可清楚看出各類別間的分佈情況, 這對本論文研究的成效探討有很大的助益。

知覺圖的分類如圖 2.5 所示, 而 MDS 即屬於非屬性為本 (non-attribute based approach) 的方法。所謂非屬性為本的方法, 是將待測點以直覺的特徵觀點, 對樣本的相似性或偏好性 (preference) 做整體的判斷, 然後在多元空間內找出待測點與樣本的定位[37]。意即不要求待測點針對預先選定的屬性進行評估, 而是讓待測點根據關注的特徵或觀點, 對樣本進行整體的相似性判斷, 以相似性判斷的量作為輸入資料, 而不需先找出相關的屬性。如此可避免受限於既定也可能不客觀的屬性, 亦不以屬性為本的方式所輸入的資料, 所得到的維度會被預設的屬性制約。因此使用

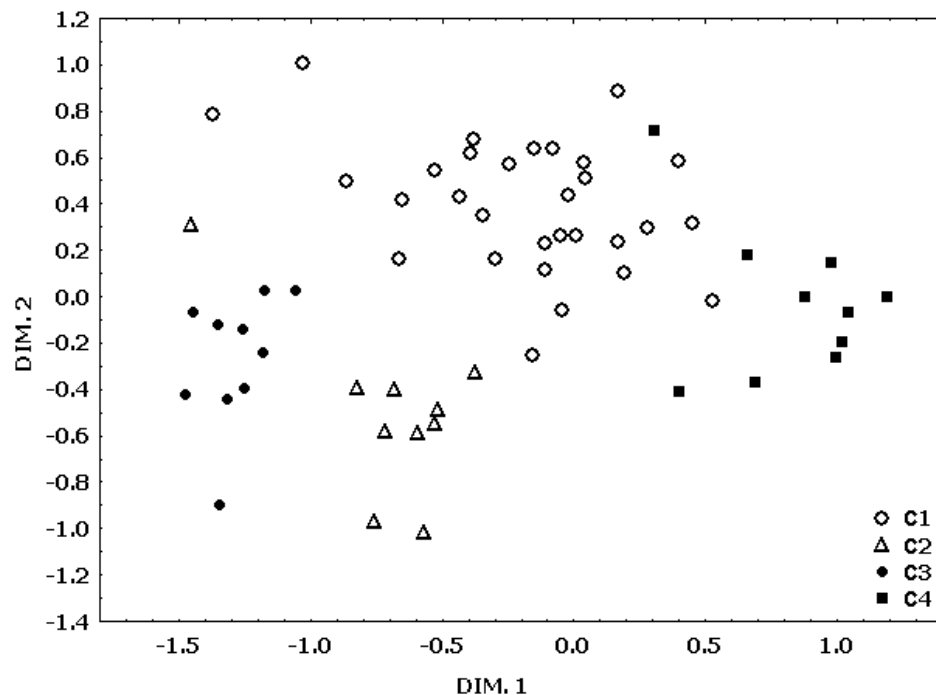


圖 2.4: 知覺圖範例

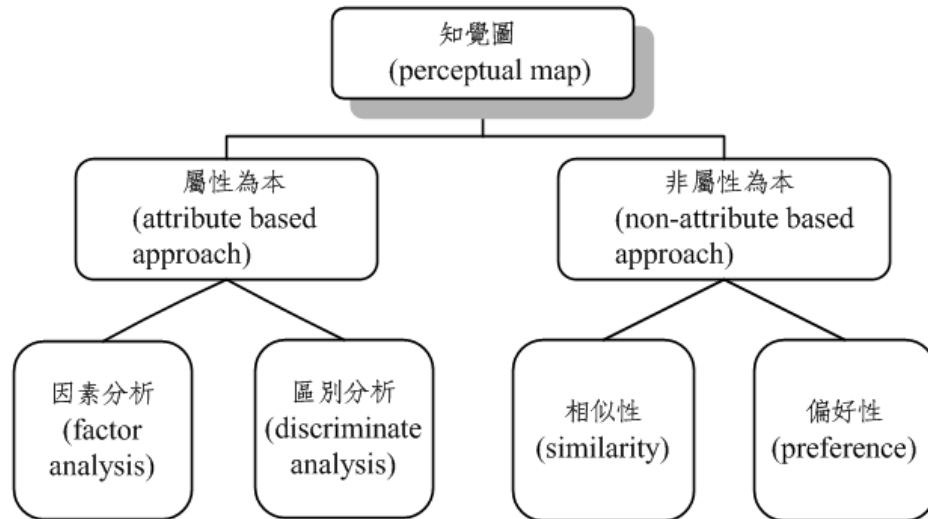


圖 2.5: 知覺圖的分類

MDS, 待測點受到研究者主觀的影響會較小, 且很可能含有更適切的結構[21, 22]。

MDS 依據待測點對於所有樣本 (或是樣本相互間) 的知覺判斷資料, 重新建構出資料所隱藏的內在結構。在資料集中, 把樣本間的相似性數值與樣本間的關係以空間的形式來描述, 就像在地圖上一樣, 描繪出樣本間的接近性 (proximity), 並將資料的結構縮減至能適切表達資料意義的最少維度。當待測點與樣本或是樣本間被評定為相似性高時, 在空間圖上的位置會較相近; 反之, 相異性高者則距離得較遠。因此, MDS 所處理的多為相似性或相異性 (dissimilarity) 資料, 而資料的相似性與相異性基本上皆假設是互斥的, 也就是資料相似性程度越高就代表相異性程度越低, 反之亦然[23]。

相似或相異性資料可為正方形對稱 (symmetric) 矩陣、正方形非對稱 (asymmetric) 矩陣、長方形 (rectangular) 矩陣等, 但為方便運算, 本論文使用正方形對稱矩陣。這些資料矩陣是由計算樣本間所有兩兩配對的相關性、相似或相異性而得, 一般而言, 採相異性資料會比相似性資料來得令人信服。進行 MDS 分析前應考量資料矩陣內的數值可否相比, 另外要注意的是在這單個矩陣所用的量尺單位必須是相同的。

2.2.2 運算機制

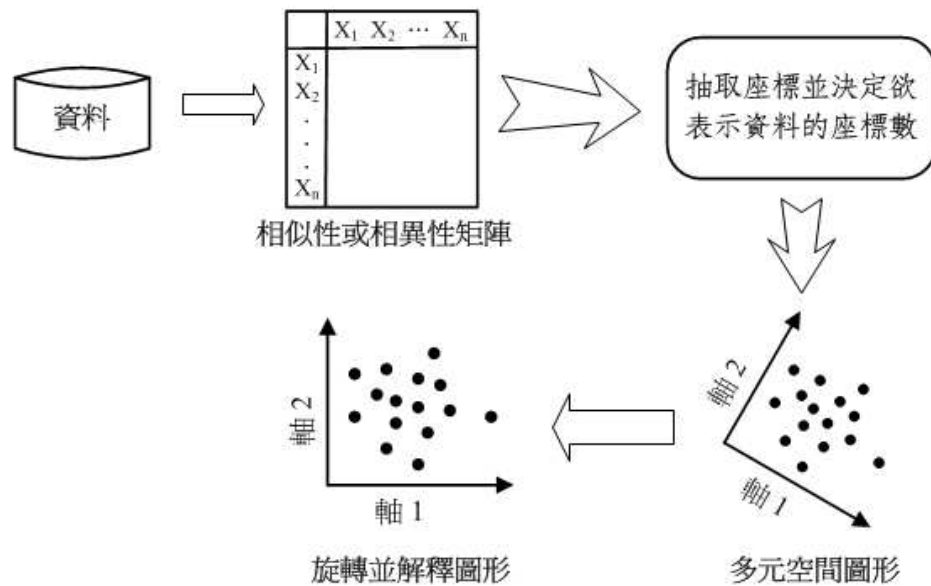


圖 2.6: 多重維度度量法流程圖

本論文是以 Torgerson (1958) 的古典多重維度度量法 (classic MDS, CMDS) 來加以改進, 屬於非加權⁵ MDS[24]。CMDS 的距離計算原理是以歐基里德距離量測法來表示二點間的相異性, 而本論文除了歐基里德距離外, 也使用動態時間扭曲法來進行運算與比較。以下的運算機制介紹將以 CMDS 為例, 本論文使用的 MDS 與 CMDS 不同處只在於輸入資料的相異性距離矩陣不同, 基本步驟如圖 2.6 所示。

參考表 2.1 的例子, 我們可以簡單的描繪出二維空間知覺圖, 如圖 2.7 所示。首先, 隨意選擇任一樣本 (本例使用 A), 並將其定點於知覺圖中央。下一個樣本 B 可定點於以 A 為圓心, B 與 A 之間的距離 (18) 為半徑的圓上。第三個樣本 C 必須滿足兩個條件: 距離 A 是 34 且距離 B 是 24, 如此將有兩種可行的架構; 而當選擇其中一種時, 剩餘的樣本定點也將隨之精確地決定。由 C 的選擇可知, 當選擇其中一種架構時, 最後產生的知覺圖其實就是另一架構產生的知覺圖作鏡射而得。而由 B 的選擇亦可知, 在圓上有無限多種架構可能, 選擇其中一種所產生的知覺圖, 其實就是其他架構所產生的知覺圖作旋轉而得。因此, 當我們隨意選擇任一架構來

⁵所謂的加權與非加權, 是根據資料矩陣的個數及分析模式來分類[25]。

	A	B	C	D	E
A					
B	18				
C	34	24			
D	31	2	25		
E	7	26	11	17	

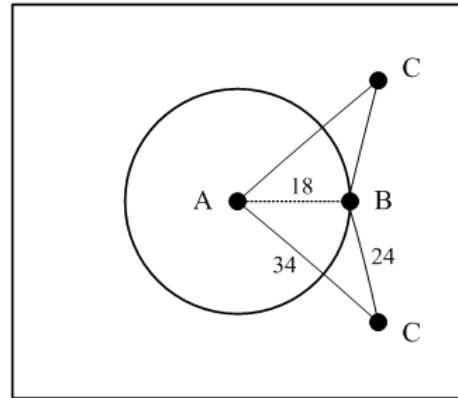


表 2.1: MDS 範例, 樣本間的距離

圖 2.7: MDS 範例, 簡易二維空間知覺圖

完成知覺圖後, 可將知覺圖再行旋轉或鏡射, 以利我們做圖形的解釋。常理上, 會旋轉圖形使得有最大變異量的軸做為第一維度, 通常顯示成在笛卡兒 (Cartesian) 平面上的東西向, 而將次大變異量的軸做為第二維度, 依此類推。

雖然上述步驟概要能輕易理解, 但在實作上其實是麻煩的, 尤其是當架構存在於二維以上的空間時。幸而有個倚賴幾何學公式的分析方法能解決這個問題, 此法與兩向量的內積 (inner product) 有關, 如圖 2.8 所示, 該二向量指的是從座標系統原點 (亦即知覺圖的零點) 至某兩點間 (在此使用 j 與 k 當例子) 的距離。因此, 兩向量的內積就是 j 、 k 兩點座標值的內積, 亦即兩向量間的餘弦 (cosine) 夾角和它們的長度作乘積。

以 CMDS 為例, 要進行 MDS 運算, 首先資料型態是為樣本間的相似性或相異性矩陣。就矩陣的元素內容值來說, 若數字越小代表兩樣本越遠離, 亦即矩陣內容值

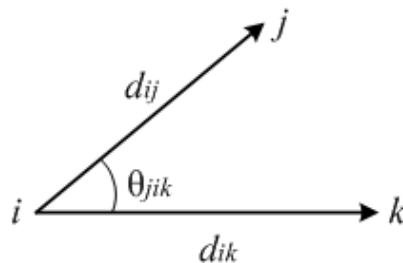


圖 2.8: 距離與內積之間的幾何關係

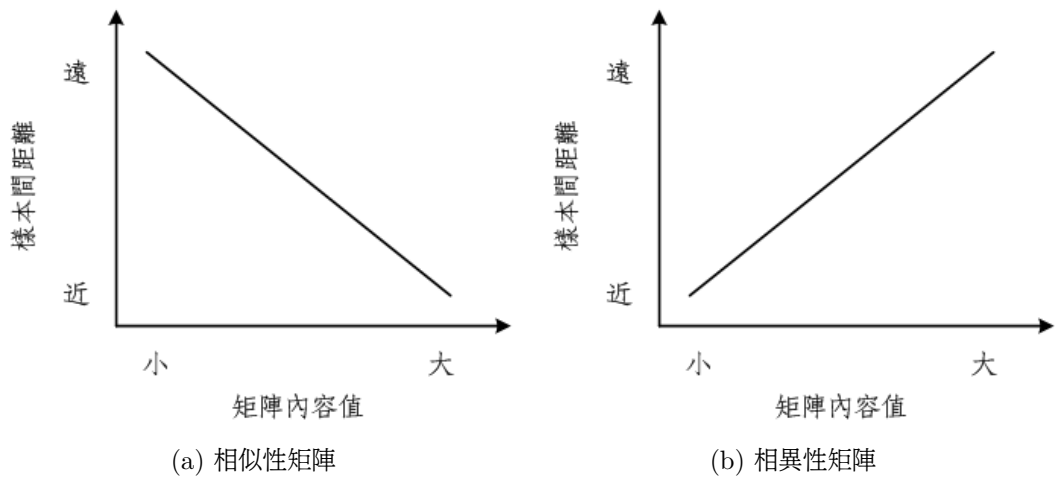


圖 2.9: 相似性與相異性矩陣的分類

與樣本間距離的線性關係之斜率為負，則為相似性矩陣；反之，若斜率為正，即為相異性矩陣，如圖 2.9 所示。我們將輸入資料的矩陣對角線元素值定義為 0 ($d_{ii} = 0$ ，對所有的 i)，且矩陣為對稱性 ($\delta_{jk} = \delta_{kj}$ ，亦即 $d_{jk} = d_{kj}$)。另外若資料是相似性的，我們先將其轉換為相異性，作法是將最大的元素內容值減去每個矩陣元素值，因此我們將輸入資料視為一個估測距離矩陣 $\mathbf{D} = [d_{ij}]$ 。

我們的運算目的是求出每個樣本的座標位置，以矩陣 $\mathbf{X} = [x_{ir}]$ 表示，其中 r 代表架構的維度數，另以 $\mathbf{x}'_i = [x_{i1}, x_{i2}, \dots, x_{ir}]$ 表示樣本 i 的座標列向量。因為任何樣本在圖上的絕對位置未定，亦即矩陣中的資訊僅足夠決定某一樣本與其餘樣本間的相關位置，因此我們可選擇任何一點作為原點。為了簡化說明，假設共有 n 個樣本，在此我們選擇以樣本 i 做為原點，再計算其餘 $(n - i)$ 個樣本在架構中與樣本 i 的相關位置。

參考圖 2.8，任兩點 j 與 k 之間的距離平方 d_{jk}^2 ，能被表示為 j 、 k 與原點間的距離 (d_{ij} 與 d_{jk}) 以及 j 、 k 之間角度 θ_{jik} 的函數[26]

$$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2 d_{ij} d_{ik} \cos \theta_{jik} \quad (2.12)$$

重整上式可得

$$-\frac{1}{2}(d_{jk}^2 - d_{ij}^2 - d_{ik}^2) = d_{ij} d_{ik} \cos \theta_{jik} \quad (2.13)$$

式 (2.13) 的等號左邊為已知項，而未知的右邊項即是內積。我們雖尚未知道如何設置 j 與 k 相互之間的關係使之滿足等式，但我們已知 d_{ij} 即為 $\|\mathbf{x}_j\|$ ，此向量長度描述點 j 與原點間的相關座標位置，而 $d_{ik} = \|\mathbf{x}_k\|$ 。另外，當兩向量的起始點固定於同一點時，則該二向量的內積為

$$\mathbf{x}'_j \mathbf{x}_k = \|\mathbf{x}_j\| \|\mathbf{x}_k\| \cos \theta_{jik} \quad (2.14)$$

因為式 (2.13) 與 (2.14) 的右邊項是相同的，故可得

$$-\frac{1}{2}(d_{jk}^2 - d_{ij}^2 - d_{ik}^2) = \mathbf{x}'_j \mathbf{x}_k \quad (2.15)$$



建立一個 $(n-1) \times (n-1)$ 的矩陣 $\mathbf{B}(i)$ ，其中 i 表示以點 i 做為原點，而 $\mathbf{B}(i)$ 中的元素為

$$b_{jk}(i) = -\frac{1}{2}(d_{jk}^2 - d_{ij}^2 - d_{ik}^2) \quad (2.16)$$

則式 (2.15) 即為

$$\mathbf{B}(i) = \mathbf{X}_i \mathbf{X}'_i \quad (2.17)$$

因為矩陣 $\mathbf{B}(i)$ 是對稱的，所以可用奇異值分解化為

$$\mathbf{B}(i) = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{U}'_i \quad (2.18)$$

其中 \mathbf{U}_i 是特徵向量矩陣，各特徵向量彼此正交，亦即 $\mathbf{U}_i \mathbf{U}_i' = \mathbf{U}_i' \mathbf{U}_i = \mathbf{I}$ 。 $\mathbf{\Lambda}_i$ 為特徵值的對角矩陣（假設所有特徵值皆為正）。故式 (2.18) 的解為

$$\mathbf{X}_i = \mathbf{U}_i \mathbf{\Lambda}_i^{1/2} \quad (2.19)$$

與 PCA 相似，由第一個特徵向量所建構的維度，將顯示出最大的變異量，亦即有著最大的特徵值，其餘特徵向量與特徵值依此類推。

MDS 的目的是要得到一個最適與最少的樣本體空間知覺圖，所有樣本根據相異或相似性呈距離關係分布，最後須合理的解釋及定義它的維度並解讀樣本間的關係。但前提是資料是否與分析後所得的樣本體空間關係的配合程度良好？該空間圖是否能以最少的構面合理解釋輸入的資料？這些都是資料投入分析後所要注意的。本研究將維度限定在 2，如此在分析最後得出的多元空間知覺圖時，將較為清楚。

2.3 k -最鄰近點分類器

2.3.1 未加權式 k -最鄰近點

k -最鄰近點 (k -Nearest Neighbors, k NN) 分類器是種基於記憶型式 (memory-based) 或是基於範例學習 (instance-based learning) 的分類方法，因此並不需要另外建立模型來實現此分類器[27]。

k NN 是利用歐基里德距離量測法選取離測試點 \mathbf{x} 最近的 k 個鄰近原型點 (prototype)，再以此 k 個原型點所屬最多的類別來判定為該待測點的類別，如圖 2.10 所示。本論文是直接將訓練資料 (training data) 視為原型點，除此之外還有各種方法可用來篩選⁶。

⁶例如 3-1 法是替每個原型點找出其最近的三個鄰居，當其中有二個以上的鄰居與該原型點的類別不同時，則將該原型點移除，不再為分類器所使用。

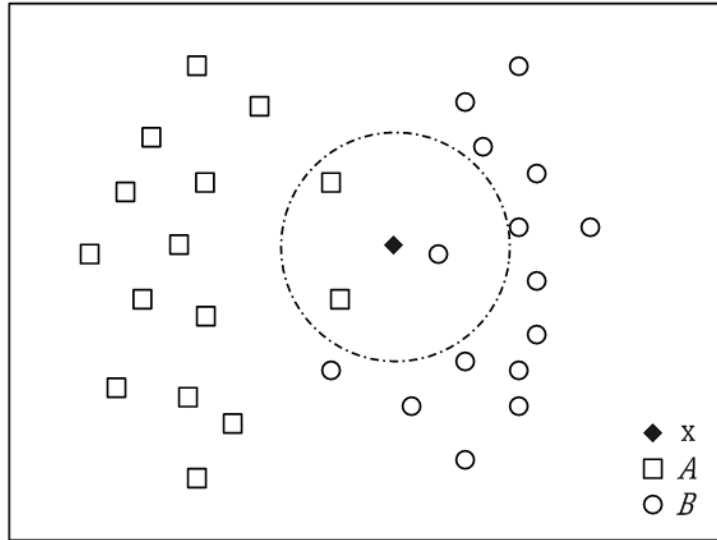


圖 2.10: k NN 分類器, $k = 3$, 則測試點 x 判定為類別 A

2.3.2 加權式 k -最鄰近點

使用 k NN 分類器時, 必須小心 k 的設定。例如若 $k = 2$, 當最近的 2 個鄰居分別屬於不同類別時, 則無法正確判斷其類別。因此, 我們使用距離加權式 k -最鄰近點法 (Distance-Weighted k NN, DW- k NN), 除了避免上述情況發生之外, 也增進了分類的準確度。

假設在樣本空間 \mathcal{R} 中有 n 個點, $\mathcal{R} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, 且共有 p 種類別, 定義第 j 種類別的加權指數 T_j 為

$$T_j = \sum_{\substack{i=1 \\ C_i=j}}^k w_i \quad (2.20)$$

其中 C_i 是離待測點 x 最近的第 i 個鄰居的類別, 而 w_i 是其權重。本論文使用的權重為

$$w_i = \frac{1}{d_i^2} \quad (2.21)$$

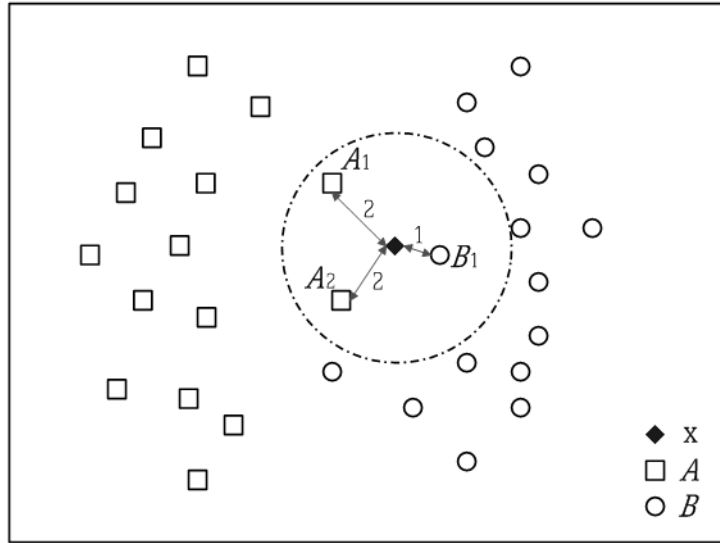


圖 2.11: DW- k NN 分類器, $k = 3$

其中 d_i 為 \mathbf{x} 與最近的第 i 個鄰居間的歐基里德距離, 則 \mathbf{x} 所屬類別定義為

$$NN_k(\mathbf{x}) = \arg \max_j T_j \quad (2.22)$$

由圖 2.11 可知, $T_A = \frac{1}{2^2} + \frac{1}{2^2} = 0.5$, $T_B = \frac{1}{1^2} = 1$, 所以 $NN_k(\mathbf{x}) = B$, 此分類結

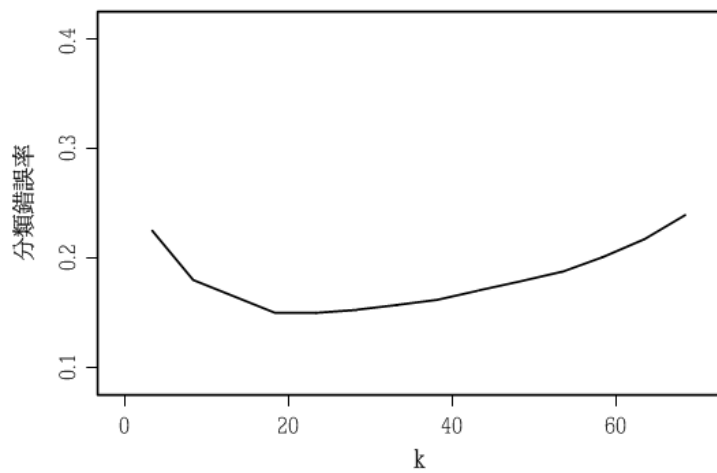


圖 2.12: k 值選擇

表 2.2: 不同訓練資料樣本數的分類成效

樣本數	k NN	DW- k NN
5%	0.64	0.69
10%	0.69	0.70
20%	0.67	0.73
40%	0.75	0.75
100%	0.78	0.78

共 26 類, 訓練資料 100% = 1190 筆, 測試資料為 496 筆。

果與圖 2.10 的 k NN 分類器不同。

k 值的選擇並沒有定, 通常是採用實驗的方式來決定, 意即驗證各種 k 值在訓練資料下的分類錯誤率, 選擇錯誤率最小的 k 值作為之後測試資料 (testing data) 所使用的 k 值。典型的 k 值錯誤率分布如圖 2.12 所示[3]。另外由表 2.2 可知, 當訓練資料樣本數越多時, 分類成效越好[38]。

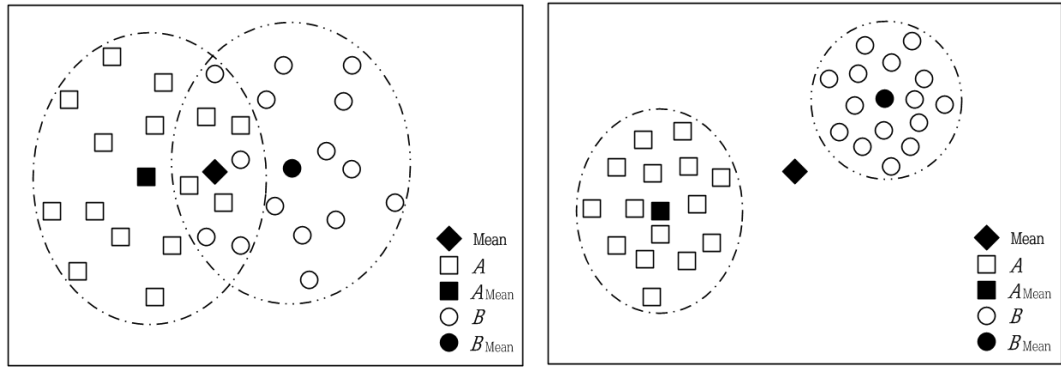
2.4 定義評估函數

在比較各種測量時間序列相似度方法的優劣時, 必須有個方法或是函式來輔助判定。本論文定義 F -試驗值 (F -test) 作為評估函數, F 值為類別間變異量 (between category variance) 與類別內變異量 (within category variance) 的比值:

$$F = \frac{S_B^2}{S_W^2} \quad (2.23)$$

其中 S_B^2 為類別間變異量, 代表各類別平均數與總平均數之間的變異量; S_W^2 為類別內變異量, 代表類別內各資料點與該類別平均數之間的變異量之總和[26, 28, 29]。

假設空間 \mathcal{R} 中有 n 個樣本點, $\mathcal{R} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, 其中 \mathbf{y}_i 在本研究中為二維空間座標向量。樣本點中共有 p 種類別, 分別稱為類別 C_j , $j = 1, \dots, p$, 則



(a) 較差的分類

(b) 較好的分類

圖 2.13: 分類範例

$$S_W^2 = \frac{1}{p} \sum_{j=1}^p \frac{\sum_{\mathbf{y}_i \in C_j} \|\mathbf{y}_i - \boldsymbol{\mu}_j\|^2}{n_j - 1} \quad (2.24)$$

$$S_B^2 = \frac{1}{p-1} \sum_{j=1}^p n_j \|\boldsymbol{\mu}_j - \boldsymbol{\mu}\|^2 \quad (2.25)$$

其中

$$\boldsymbol{\mu}_j = \frac{1}{n_j} \sum_{\mathbf{y}_i \in C_j} \mathbf{y}_i \quad (2.26)$$

$$\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i = \frac{1}{n} \sum_{j=1}^p n_j \boldsymbol{\mu}_j \quad (2.27)$$

n_j 代表類別 C_j 裡的樣本數, $\boldsymbol{\mu}_j$ 代表所有在類別 C_j 裡的樣本之平均值 (亦即各類別樣本之中心座標值), 而 $\boldsymbol{\mu}$ 代表所有樣本之平均值。

圖 2.13 中, \blacklozenge 代表 $\boldsymbol{\mu}$ 、 \square 代表類別 C_A 中的樣本、 \blacksquare 代表 $\boldsymbol{\mu}_A$ 、 \circ 代表類別 C_B 中的樣本、 \bullet 代表 $\boldsymbol{\mu}_B$ 。由圖可知, 若希望有最好的分類成效, 則應該使各類別間儘可能相距越遠越好, 而各類別內則是聚的越密越好; 換言之, 類別內之異質性很小, 而類別間之異質性很大, 意即希望 S_B^2 越大, 而 S_W^2 越小越好。另由式 (2.23)

至 (2.25) 可知, 當 F 值越大, 意即當各類別之平均數相差越大, 而各類別之標準差越小時, 則表示分類成效越好。而當類別間變異量大於類別內變異量時, 即可說明總體變異量大部分是由各類別間的差異所造成。



第 3 章

時間序列之相似度量測

3.1 歐基里德距離量測法

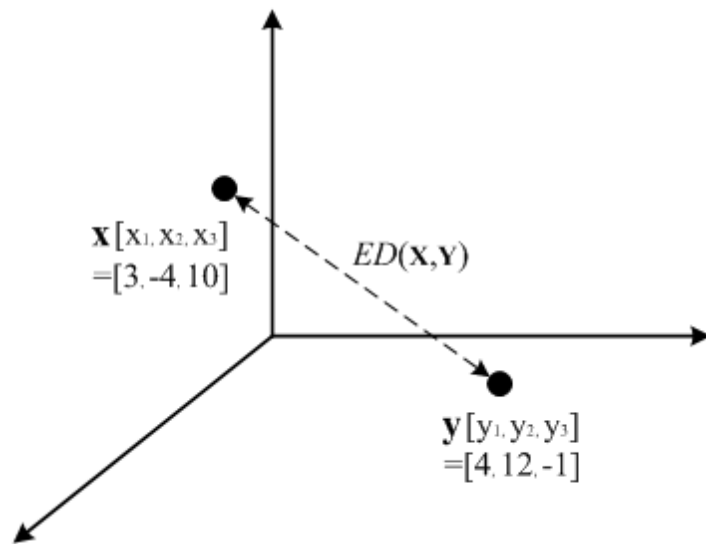


圖 3.1: 歐基里德距離量測法範例

傳統上在衡量兩測量點之間的距離時，較常使用歐基里德距離量測法 (Euclidean Distance Metric, EDM) 來表示，因其較為直觀且易於計算。本論文中，可將各時間序列樣本視為在一 n 維歐基里德空間中的一個點 (n 為時間序列的長度)，如此即可量測兩點間歐基里德距離。假設在 n 維輸入空間中，有 m 個樣本數的資料 $\mathbf{X}_{m \times n} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]'$ ，其中 $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]$ 表示第 i 個樣本。則目標點 $\mathbf{x}_0 = [x_{01}, x_{02}, \dots, x_{0n}]$ 與樣本點 \mathbf{x}_i 的歐基里德距離 $ED(\mathbf{x}_0, \mathbf{x}_i)$ 為

$$ED(\mathbf{x}_0, \mathbf{x}_i) = \|\mathbf{x}_0 - \mathbf{x}_i\| = \sqrt{\sum_{j=1}^n (x_{0j} - x_{ij})^2} \quad (3.1)$$

以圖 3.1 為例, \mathbf{x} 與 \mathbf{y} 間的歐基里德距離 $ED(\mathbf{x}, \mathbf{y})$ 為 $\sqrt{1^2 + (-16)^2 + 11^2} = \sqrt{378} \simeq 19.44$ 。EDM 雖有著易於計算的優點, 但其缺點為不允許不相同的基準線 (baseline) 以及不同的尺度比率 (scale)。例如, 股票 \mathbf{x} 在 \$100 上下作波動, 而股票 \mathbf{y} 卻在 \$40 上下作波動; 或是股票 \mathbf{x} 在 \$95 與 \$115 間做波動, 而股票 \mathbf{y} 卻在 \$20 與 \$30 間做波動, 這些情況若使用 EDM 都是不適當的[30]。

3.2 動態時間扭曲法

3.2.1 特性簡介

在上節中, 我們可以得知 EDM 在計算兩時間序列的方式上, 是將同時時間點上的兩筆資料作距離運算, 最後加總整段時間序列的距離, 依此作為相似度的參考。不過現實例子中, 我們常可發現兩段時間序列間有時只是做了位移的關係, 整體上以人眼看起來是非常相似的, 但若以 EDM 的單一時間點上資料距離來看, 卻會差距很大, 也因此造成了最後呈現出兩時間序列極不相似的運算結果, 而與事實不符。

以圖 3.2 為例, 就人眼看來, 編號第 1、2、3 號應該是非常相似的, 且與編號 4

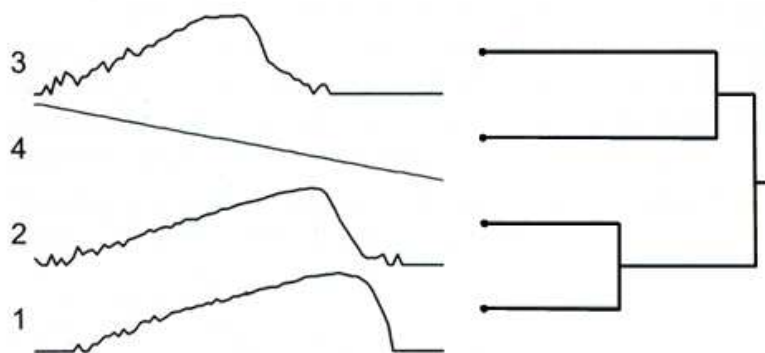


圖 3.2: 歐基里德距離量測法相似度分類範例

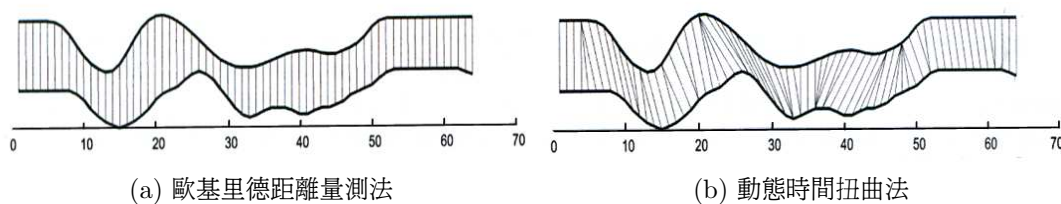


圖 3.3: 相似度運算圖解

有極大的不同, 但若以 EDM 來做相似度運算的話, 就會呈現圖中錯誤的分類結果。因此, 我們必須使用一種能判斷兩時間序列整體趨勢 (trend) 的方法來作為相似度的依據。

本論文選擇了在語音辨識問題中常使用的動態時間扭曲法 (Dynamic Time Warping, DTW) 作為另一種相似度量測的依據, 該法是以 EDM 為基礎而加以改進, 主要原理是允許時間序列能重複 (或延伸) 其中的元素 (亦即單一時間點上的資料)。DTW 與 EDM 的運算基本概念不同處如圖 3.3 所示, 圖中兩點間的連線代表著歐基里德距離運算, 但 DTW 與 EDM 不同的是, 它可以允許計算兩個不同時間點上的資料, 如此即可達成趨勢判斷的目的。以圖 3.2 的例子, 改採 DTW 做運算後所得的相似度分類結果如圖 3.4 所示, 可發現分類成效比 EDM 好很多[31, 32, 33]。

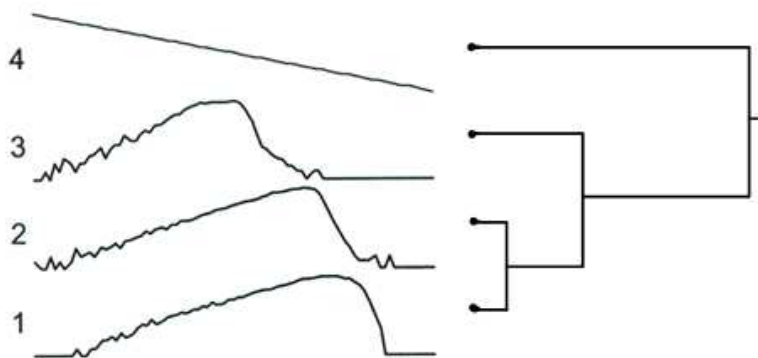


圖 3.4: 動態時間扭曲法相似度分類範例

3.2.2 視窗式動態時間扭曲法

假設兩時間序列 \mathbf{q}, \mathbf{c} , 各為長度 n 與 m

$$\begin{aligned} \mathbf{q} &= [q_1, q_2, \dots, q_i, \dots, q_n] \\ \mathbf{c} &= [c_1, c_2, \dots, c_j, \dots, c_m] \end{aligned} \quad (3.2)$$

接下來建構一 $n \times m$ 的距離矩陣, 其中第 (i, j) 個元素內容為兩時間序列的點 q_i 與 c_j 之間的歐基里德距離。圖 3.5 顯示一扭曲路徑 (warping path) 範例, 扭曲路徑代表的即是圖 3.3 (b) 中, 兩時間序列間的點與點配對過程。扭曲路徑 \mathcal{W} 為距離矩陣元素間的連續集合, 其中第 k 個元素定義為 $w_k = (i, j)_k$, 因此

$$\mathcal{W} = \{w_1, w_2, \dots, w_k, \dots, w_K\} \quad \max(m, n) \leq K < m + n - 1 \quad (3.3)$$

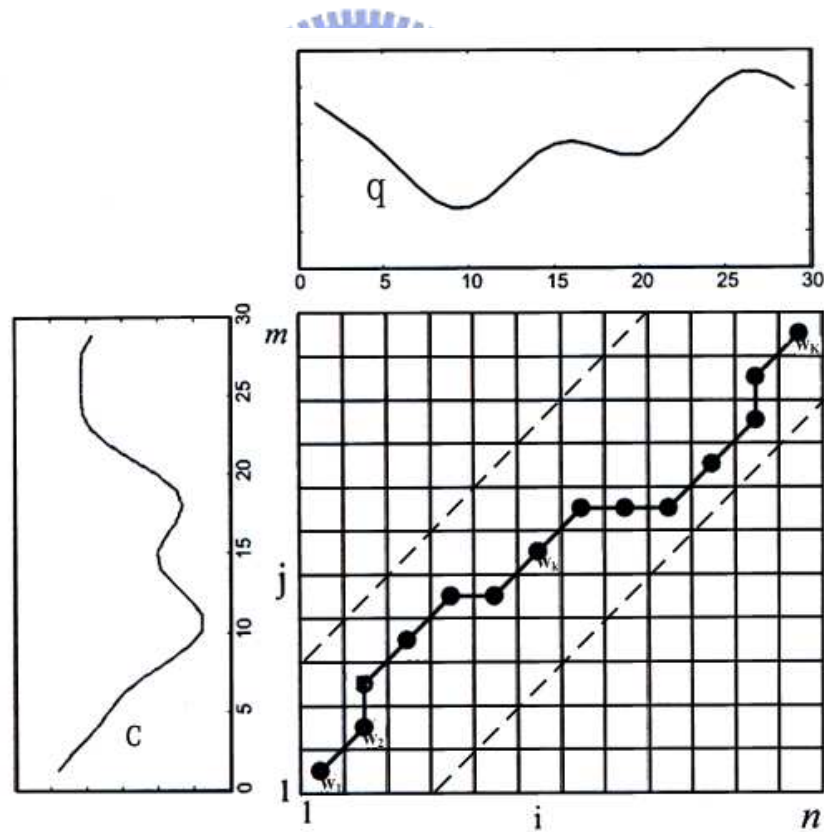


圖 3.5: 動態時間扭曲法路徑範例

其中

$$w_k = ED(q_i, c_j) = \sqrt{(q_i - c_j)^2} \quad (3.4)$$

而兩時間序列的相似度 $DTW(\mathbf{q}, \mathbf{c})$ 即可用 \mathcal{W} 來定義, 亦即

$$DTW(\mathbf{q}, \mathbf{c}) = \min\left(\frac{\sum_{k=1}^K w_k}{K}\right) \quad (3.5)$$

數值越小代表越相似。由於扭曲路徑不只一條, 因此式 (3.5) 中採用最小值來做最後的決定路徑; 而每條扭曲路徑可能會有不同的長度, 因此分母的 K 是用來做補償運算[34, 35]。

扭曲路徑受到下列幾點限制:

- 邊界條件 (boundary conditions)

$w_1 = (1, 1)$ 、 $w_K = (n, m)$, 亦即扭曲路徑必須開始與結束在距離矩陣的對角元素。以圖 3.5 為例, 即是由距離矩陣的左下角元素開始, 而在右上角結束。

- 單調性 (monotonicity)

若給定 $w_k = (a, b)$, 而 $w_{k-1} = (a', b')$, 則 $a - a' \geq 0$ 且 $b - b' \geq 0$ 。換句話說, 以圖 3.5 的例子而言, 扭曲路徑將不會往下或往左走。

- 連續性 (continuity)

若給定 $w_k = (a, b)$, 而 $w_{k-1} = (a', b')$, 則 $a - a' \leq 1$ 且 $b - b' \leq 1$, 亦即時時間序列中沒有元素會被跳過。以圖 3.5 為例, 此連續性限制將使扭曲路徑前進至鄰近的距離矩陣元素 (包含斜角鄰近元素)。

除了上述扭曲路徑的限制之外, 我們亦可多加上一扭曲視窗 (warping window) 的條件, 本論文將使用扭曲視窗的 DTW 稱為視窗式動態時間扭曲法。所謂的扭曲視窗, 可視做為扭曲路徑的最大範圍。圖 3.5 中的兩條虛線 $j = i + W$ 、 $j = i - W$ (W 為扭曲視窗大小) 即代表著視窗範圍, 扭曲路徑不得超過此範圍, 其中 $W = 3$ 。

3.2.3 最小累進距離式動態時間扭曲法

由於視窗式動態時間扭曲法的扭曲路徑不只一條，必須找出所有可能性後，再採用最小距離值的路徑，但這會造成計算上的繁瑣。在此亦使用另一種計算較為有效率的 DTW，在本論文中稱之為最小累進距離式動態時間扭曲法，該法是利用遞迴的累計距離運算來實現。定義距離矩陣的第 (i, j) 元素內容為 $d(i, j)$ (即 q_i 與 c_j 間的歐基里德距離)，而扭曲路徑行進到第 (i, j) 元素時的累進距離為 $\gamma(i, j)$ ，則

$$\gamma(i, j) = d(i, j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \quad (3.6)$$

當 \mathcal{W} 中的第 k 個元素 $w_k = (i, j)_k$ 受到 $i = j = k$ 的限制時，此時的 DTW 即等同歐基里德距離量測法，但前提是兩時間序列需有相同長度[34, 35]。



第 4 章

實驗與討論

4.1 資料產生與實驗流程

4.1.1 資料產生

本研究所使用的資料是自行模擬產生，並非實際取自業界的資料。回顧第一章的問題描述，除了正常狀態¹的資料外，尚有五種不同類型的異常資料需要被分類²，因此我們得先產生此六種類別的時間序列資料集。圖 4.1 展示本論文所使用的模擬資料其產生流程，由圖可知資料的產生全建立在正常資料樣板 (template) 之上，因為本研究除了偵測正常工作範圍內的異常訊號，也希望能對偵測出的異常訊號加以分類，以藉此判定異常的發生情況與解決方式。

本研究所定義的正常資料樣板如圖 4.2 (a) 所示，樣板長度為 150；依此樣板為基底，並允許四個轉折點做些許的左右移動 (± 3 以內，此舉是為了增加正常資料之時間序列的變化性)，在此新樣板上加上隨機亂數 (± 10 以內) 以產生正常資料之時間序列，如此重複操作直至產生所需的足夠時間序列數量，如圖 4.2 (b) 所示。所需的正常資料之時間序列產生後，接下來即得設定出正常狀態下的工作範圍，本論文使用資料集的均值加上 ± 3 個標準差作為正常工作範圍的依據。假設正常時間序列集為 $\{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n\}$ ，則正常工作範圍 Range_{\max} 與 Range_{\min} 為

¹所謂的正常狀態，即指工業程序操作上，不會發生問題的工作狀態。

²本論文所稱的正常資料，意指工業程序操作過程或結果沒有問題時的資料；反之，本論文所稱的異常資料，即指操作過程或結果有問題時的資料。

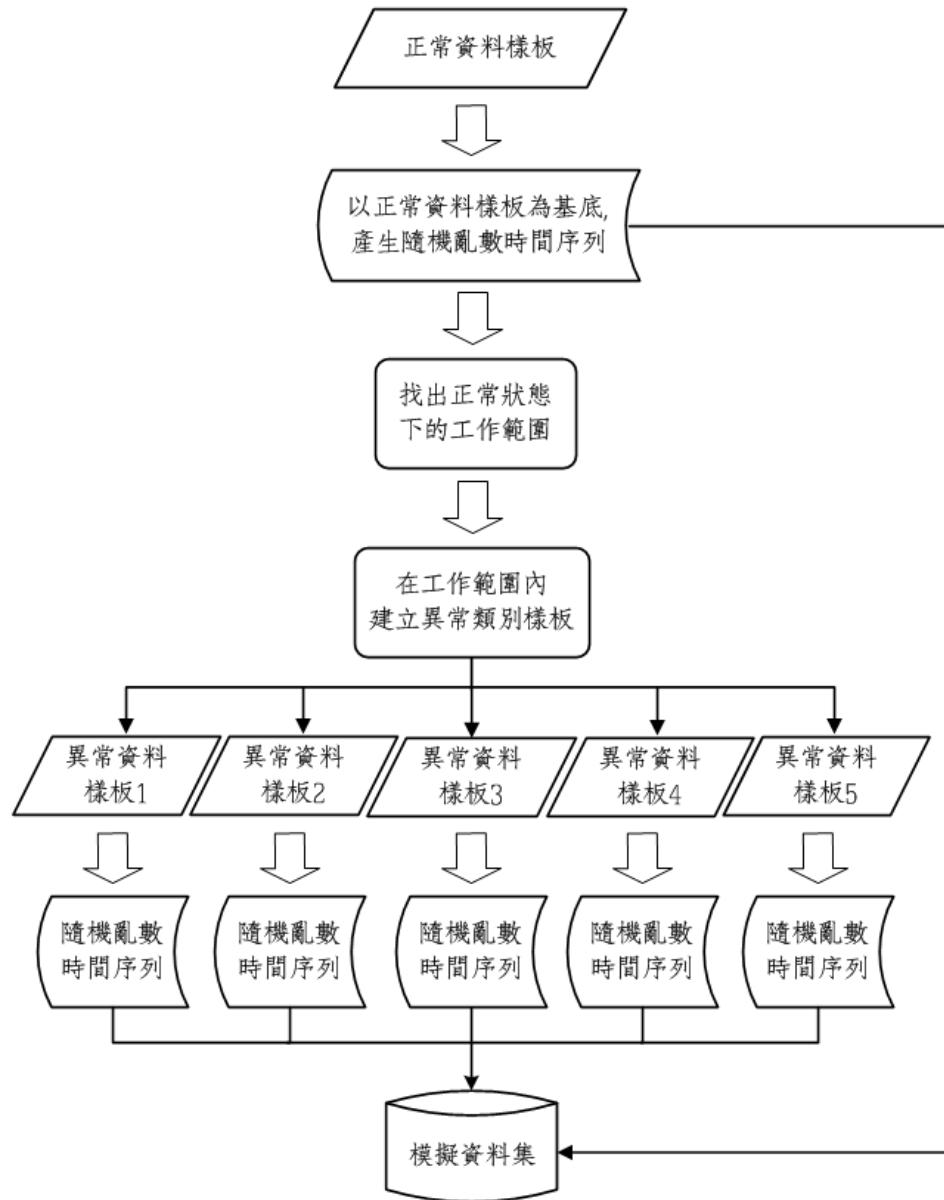


圖 4.1: 資料產生流程圖

$$\begin{aligned}
 \text{Range}_{\max}(i) &= \bar{g}_i + 3 \sqrt{\frac{1}{n-1} \sum_{j=1}^n (g_{ji} - \bar{g}_i)^2} \\
 \text{Range}_{\min}(i) &= \bar{g}_i - 3 \sqrt{\frac{1}{n-1} \sum_{j=1}^n (g_{ji} - \bar{g}_i)^2}
 \end{aligned} \tag{4.1}$$

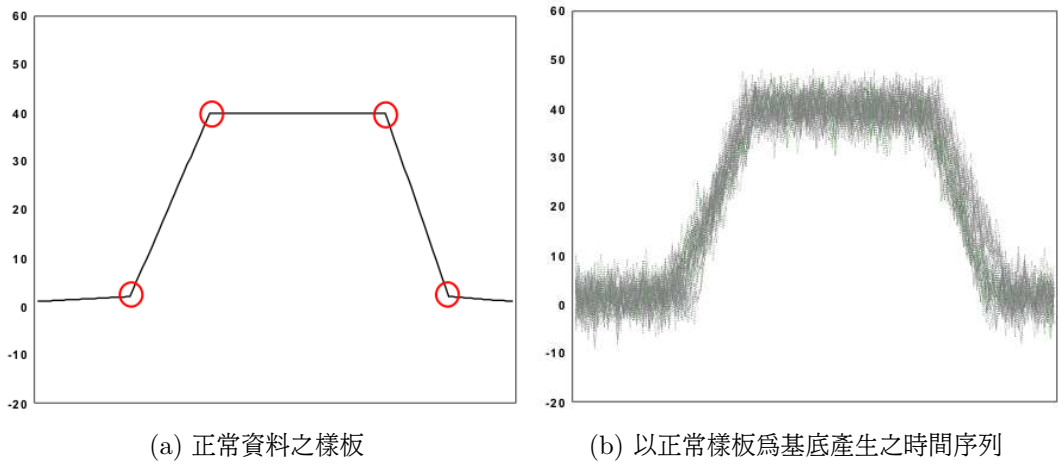


圖 4.2: 正常資料時間序列之產生

其中 i 表示第 i 個時間點, $i = 1, 2, \dots, 150$ 。 g_{ji} 代表 g_j 的第 i 個時間點資料, 而 \bar{g}_i 代表第 i 個時間點的時間序列資料集平均值, 意即

$$\bar{g}_i = \frac{1}{n} \sum_{j=1}^n g_{ji} \quad (4.2)$$

以圖 4.2 為例, 所定義出的正常工作範圍如圖 4.3 所示。

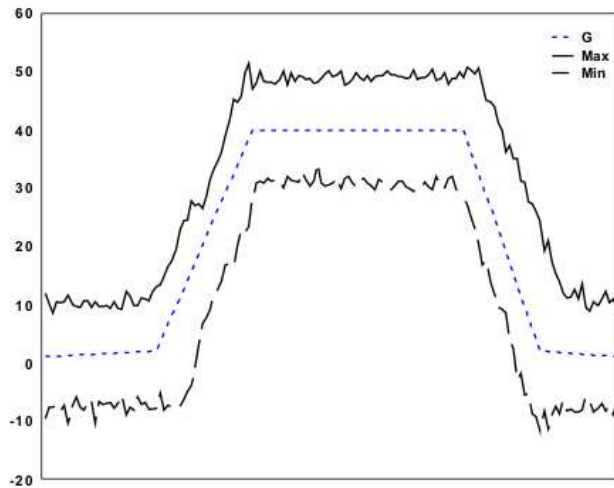
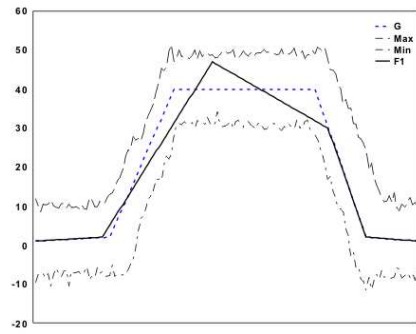
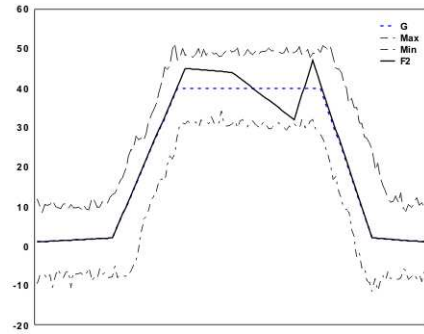


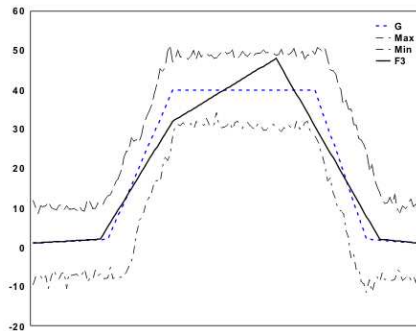
圖 4.3: 正常狀態之工作範圍



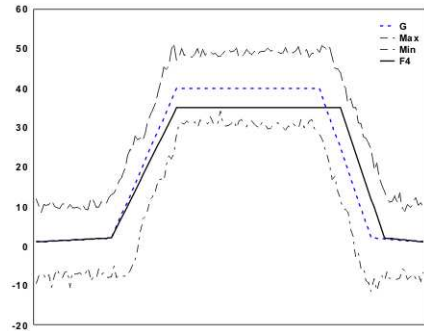
(a) 異常類別樣板1



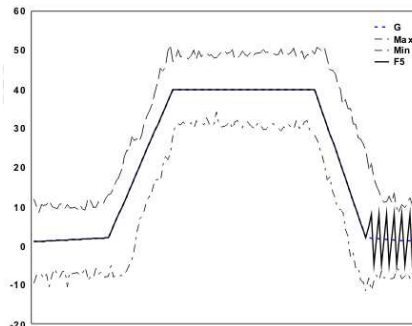
(b) 異常類別樣板2



(c) 異常類別樣板3



(d) 異常類別樣板4



(e) 異常類別樣板5

圖 4.4: 異常類別資料樣板

工作範圍決定好後，接下來將產生另外五種異常資料類別的時間序列。與正常資料製作方式相同，我們必須先產生出樣板，再依此產生隨機的時間序列資料。因為本研究的目標是希望在正常工作範圍下，仍能發現異常訊號的存在並加以分類，因此本研究所產生的隨機時間序列不管異常與否皆需在正常工作範圍內。各異常類別樣板如圖 4.4 所示，其中正常範圍內的虛線是正常資料的樣板，各異常類別在實際業界資料所代表的意義在第一章中已介紹，本研究為了儘可能測試系統模型的成

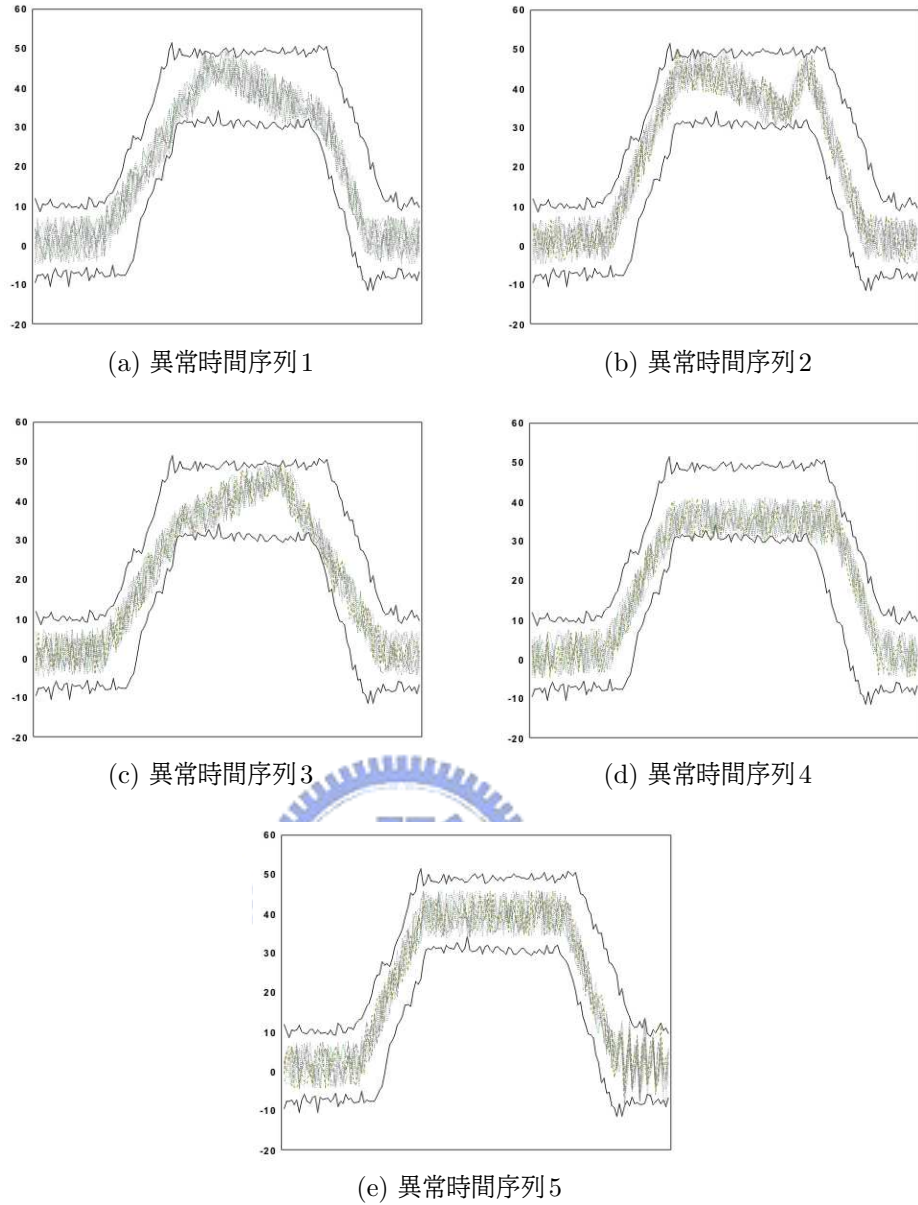


圖 4.5: 異常類別資料之時間序列

效，因此實驗設計異常類別的樣板越近似正常的樣板越好，由圖 4.4 可觀察之。與正常樣版不同的是，異常樣板並不會將轉折點做些需左右移動，因此產生的各異常類別樣板就僅此一個，此舉除了避免異常樣板跑出正常工作範圍之外，亦可遵照實驗設計的原始理念——異常樣板不要偏離正常樣板太多。之後即可以此五種樣板為基底，加上隨機亂數產生出所需的各異常類別時間序列。另外必須注意的是，加上的隨機亂數不能使時間序列超出正常工作範圍，圖 4.5 是以圖 4.4 的樣板產生出的

時間序列。在各實驗中，五種異常類別所使用的時間序列數量皆一致，如此在分類時就不需考慮到各類別資料數量不同的問題。至於加在樣板上的隨機亂數範圍對實驗結果的影響，將在之後的實驗中加以討論。

4.1.2 實驗流程

本論文實驗的整體流程如圖 4.6 所示，其中第一步的模擬資料建立在節 4.1.1 中已詳細說明，接下來我們必須處理原始的時間序列資料集，得出系統實際所需的輸入資料——相異性距離矩陣。此矩陣為一方陣，乃將每個時間序列與其他時間序列間的相似度以相異性距離量測數據來表示，亦即若整個時間序列資料集（含正常與異常類別資料）有 n 筆資料，則會得出一 $n \times n$ 的相異性距離矩陣，其中各時間序列與自身的距離定義為 0，因此該矩陣中的元素值越小代表兩時間序列越相似，反之則越不相像，此即“相異性”的定義。至於相似度距離量測的方式，本研究在各實驗中皆探討三種方式的成效：歐基里德距離量測法、視窗式動態時間扭曲法（視窗大小為另一將探討的可變參數）與最小累進距離式動態時間扭曲法，因此實驗流程中產生相異性距離矩陣的這部份，將會有六個³矩陣產生。

系統所需要的輸入資料準備好之後，即可進行實驗階段。本研究主要包括兩個實驗，類別內與類別間的不同相似度距離量測方式之分類成效探討，與時間序列的位移量與雜訊大小對分類成效的影響，詳細的實驗步驟與結果將在節 4.2 與 4.3 中說明。

參考圖 4.6 可知，實驗所得出的輸出資料為二維空間知覺圖，更正確的說，實驗的輸出資料為各時間序列在知覺圖中的座標值。知覺圖可幫助我們直覺地看出各類別的分類狀況，但若要有實際數據來表示分類的成效，則需利用座標值來計算我們定義的分類成效評估函數—— F -試驗值。總的來說，此 F 值越大代表分類的成效越好，但對類別內實驗而言，我們希望的是同一類別內的時間序列雖然有著位移或是雜訊等不同狀況發生，仍不會因此被分成多種不同類別；因此在該實驗中， F 值

³因為視窗式動態時間扭曲法的視窗大小為一可變參數，因此實驗中產生視窗大小 $W = \{5, 10, 15, 20\}$ 的四個相異性距離矩陣，加上歐基里德距離量測法與最小累進距離式動態時間扭曲法，就共有六個相異性距離矩陣供實驗進行與討論。

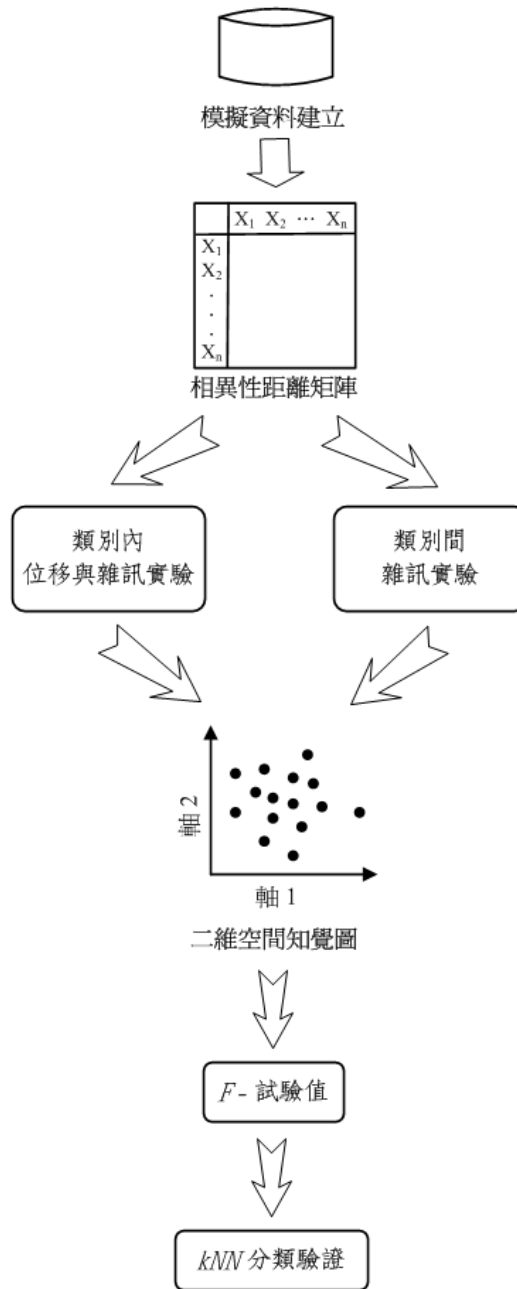


圖 4.6: 整體實驗流程圖

越小反而代表實驗成效越好，詳細討論將在節 4.2 中說明。

當上述的“實”作部分完成，還需要“驗”證不同方式的實驗結果與分類成效好壞，本研究使用的是去一法 (leave one out)，將原有模擬資料集內的各時間序列逐一當作待測資料，其餘的 $n - 1$ 個時間序列作為樣本空間，再使用加權式 k -最鄰

近點分類法來分類該待測資料屬於正常或是異常的資料類別？若為異常類別又是屬於何種異常 (包含正常狀態, 共有六種可分類類別)? 由於使用的是建構系統模型的原始資料, 因此我們已知該資料實際上是屬於何種類別, 再對照系統分類出來的類別, 即可判定該系統的分類錯誤率⁴, 相關討論詳見節 4.4。

4.2 類別內實驗

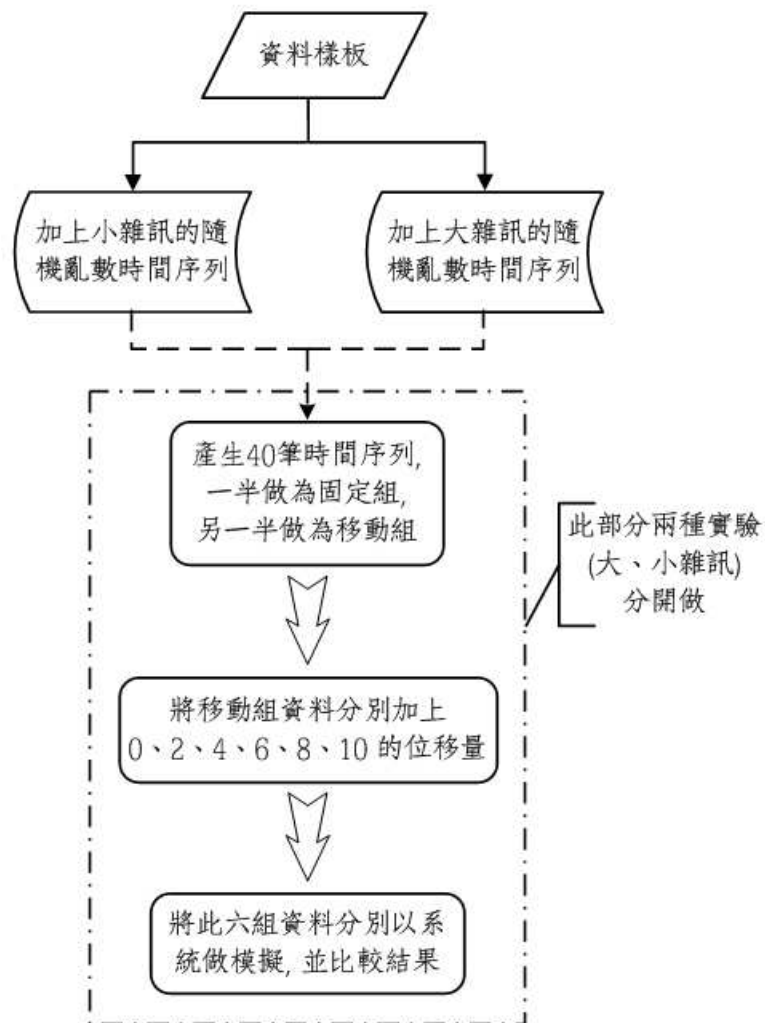


圖 4.7: 類別內實驗流程圖

⁴習慣上, 評定系統模型的好壞是以錯誤率 E 為依據, 而不使用正確率 T , 雖然兩者有著 $E+T=1$ 的關係。

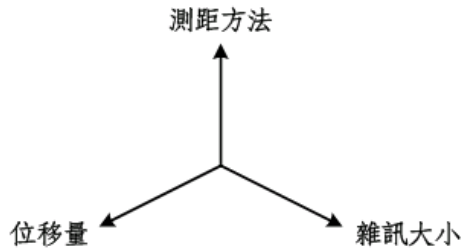


圖 4.8: 類別內實驗之控制變數

類別內 (within category) 實驗流程如圖 4.7 所示, 本論文在此單一類別實驗是使用正常資料樣板為基底, 並將此實驗分為小雜訊與大雜訊兩部份。如圖 4.8 所示, 本實驗之控制變數有三種, 除了雜訊大小之外, 另有時間序列位移量以及相似度距離量測方法兩種因素待討論。

以小雜訊實驗為例, 我們在正常資料樣板上加上 ± 2 以內的隨機亂數 (此即雜訊), 先產生出 40 筆時間序列, 其中 20 筆設為固定組資料, 亦即之後的各組實驗皆不會更動到此 20 筆資料, 而剩餘的 20 筆資料則設為移動組資料, 該組資料將在之後用來實驗不同位移量對分類結果的影響。本論文測試了時間序列延後 (delay)

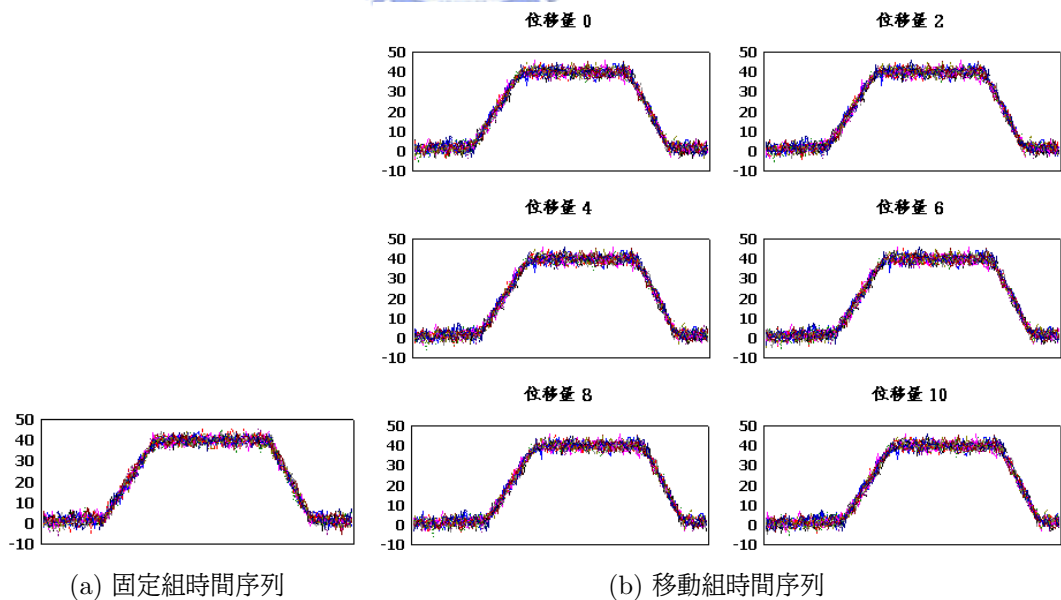


圖 4.9: 類別內實驗之小雜訊時間序列

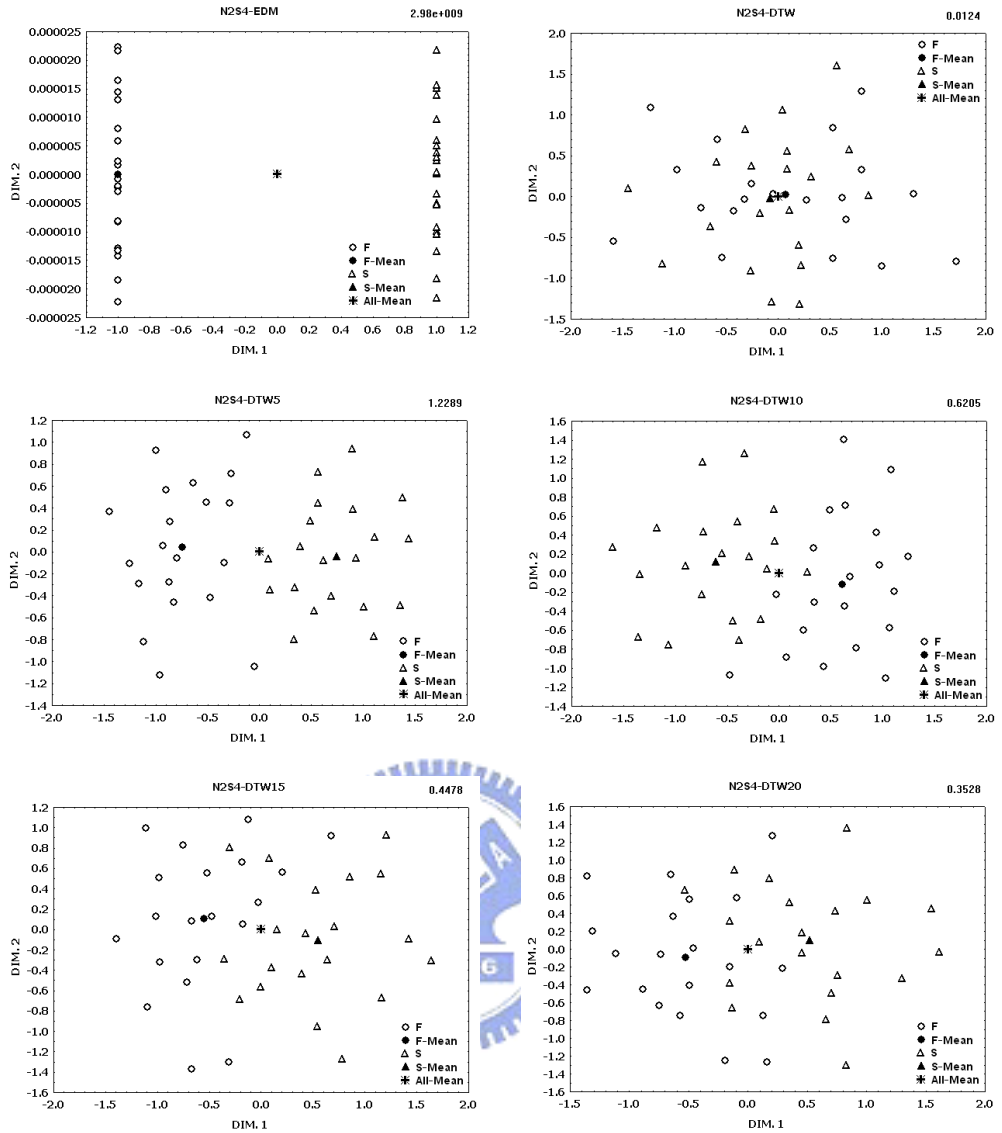


圖 4.10: N2S4之知覺圖

位移 (亦即時間序列向右移動) {2、4、6、8、10} 等五種時間, 加上未位移共六組數據。圖 4.9 展示了此六組數據, 可將資料分為固定組與移動組兩類, 而我們的目標是找出一個較好的方法, 使其能將兩類的資料視為一類。因為此兩類資料本為同一樣板所產生, 雖有位移的情況發生, 仍該歸屬為同一類, 因此最後得出的 F -試驗值應該要越小越好。

圖 4.10 展示了部份的實驗結果知覺圖, 其中 N2S4 代表 “雜訊為 ± 2 以內, 位移量為 4”, 而各圖右上角的數值即為 F -試驗值。各圖上的標題, EDM 代表使用歐

基里德距離量測法, DTW 代表使用最小累進距離式動態時間扭曲法, DTW5 代表使用視窗大小為 5 的動態時間扭曲法, 其餘依此類推。

由圖 4.10 觀察可知, 當固定組 (F) 與移動組 (S) 兩組數據分的越開時, F -試驗值也就越大, 代表著系統將此二組資料視為不同的兩個類別。從知覺圖的分佈與 F -試驗值可得知, 使用 EDM 的系統較無法判別位移後的時間序列仍屬於同一類, 而 DTW 的視窗大小也決定系統判別的成效, 當視窗值設定越大, F -試驗值就越小。由於最小累進距離式動態時間扭曲法並不需設定視窗大小, 意即視窗可視為任意長度, 因此使用該法的系統成效會較 DTW20 來的更好。另外知覺圖中也顯示了各類別 (F-Mean, S-Mean) 以及整體資料 (All-Mean) 的均值, 此均值為座標值的均值, 因此可發現全體均值在各圖中皆定位在知覺圖的原點 (0,0) 附近。

表 4.1 列出了類別內小雜訊實驗的所有 F -試驗值數據, 圖 4.11 則展示各相似度量測方法的 F -試驗值趨勢, 由於有些 F -試驗值過大, 為方便圖形顯示與判讀, 故將 F 值取對數 (log)。由圖可看出各相似度量測方法的 F -試驗值趨勢皆隨著位移量的增加而變大, 且 EDM 與 DTW5 的上升趨勢最為顯著。在此實驗中, 我們期望 F 值能越小越好, 且希望 F 值曲線能盡可能平穩, 如此即可代表該量測方法對位移量的增加較不敏感, 也就可用該量測法來判定時間序列的相似度; 意即當時間序列只是有著位移而整體趨勢是差不多的時候, 該方法不會將之分為不同的類別。由表與圖可得知, 各方法的優劣度為: $DTW > DTW20 > DTW15 > DTW10 > DTW5 > EDM$ 。另外注意 DTW5 的曲線, 可發現在位移量為 6 時即急速攀升,

表 4.1: N2 與不同位移量下的各系統 F -試驗值

位移量	EDM	DTW	DTW5	DTW10	DTW15	DTW20
0	0.0204	0.0061	0.0037	0.0027	0.0137	0.0270
2	3.8095	0.0085	0.6729	0.3312	0.3003	0.2009
4	2.98e+09	0.0124	1.2289	0.6205	0.4478	0.3528
6	3.01e+09	0.0256	5.8745	0.7619	0.4592	0.3918
8	8.24e+09	0.0351	4.80e+09	1.3305	0.7962	0.6702
10	8.58e+09	0.1165	4.85e+09	1.5347	1.0065	0.8529

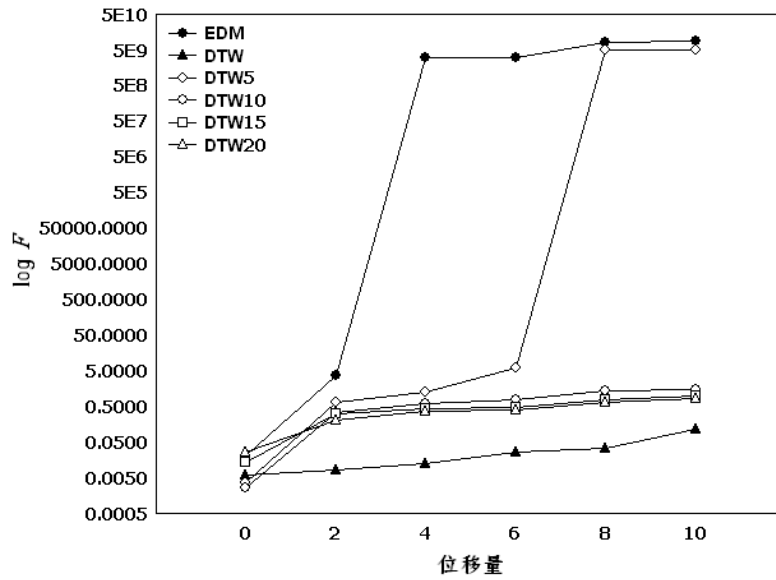


圖 4.11: N2與不同位移量下的各系統成效

這是因為位移量在此時已超過了視窗所能掃描的長度, 因此才會判定此兩組實驗數據為不同的類別; 而其餘的 DTW 量測方法因位移量最大為 10, 並未超過其視窗長度, 因此曲線看來較平穩。

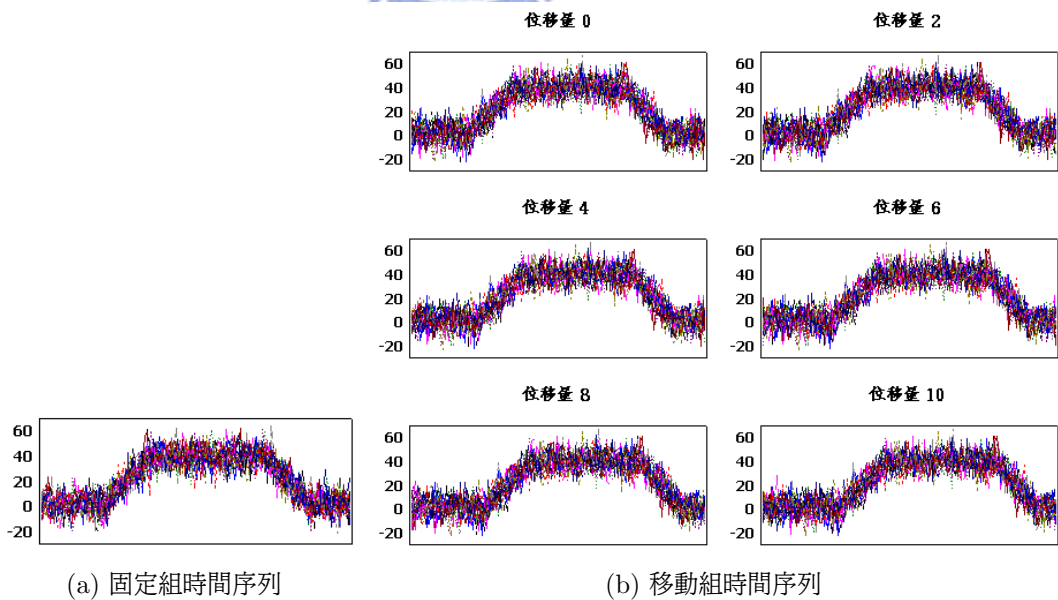


圖 4.12: 類別內實驗之大雜訊時間序列

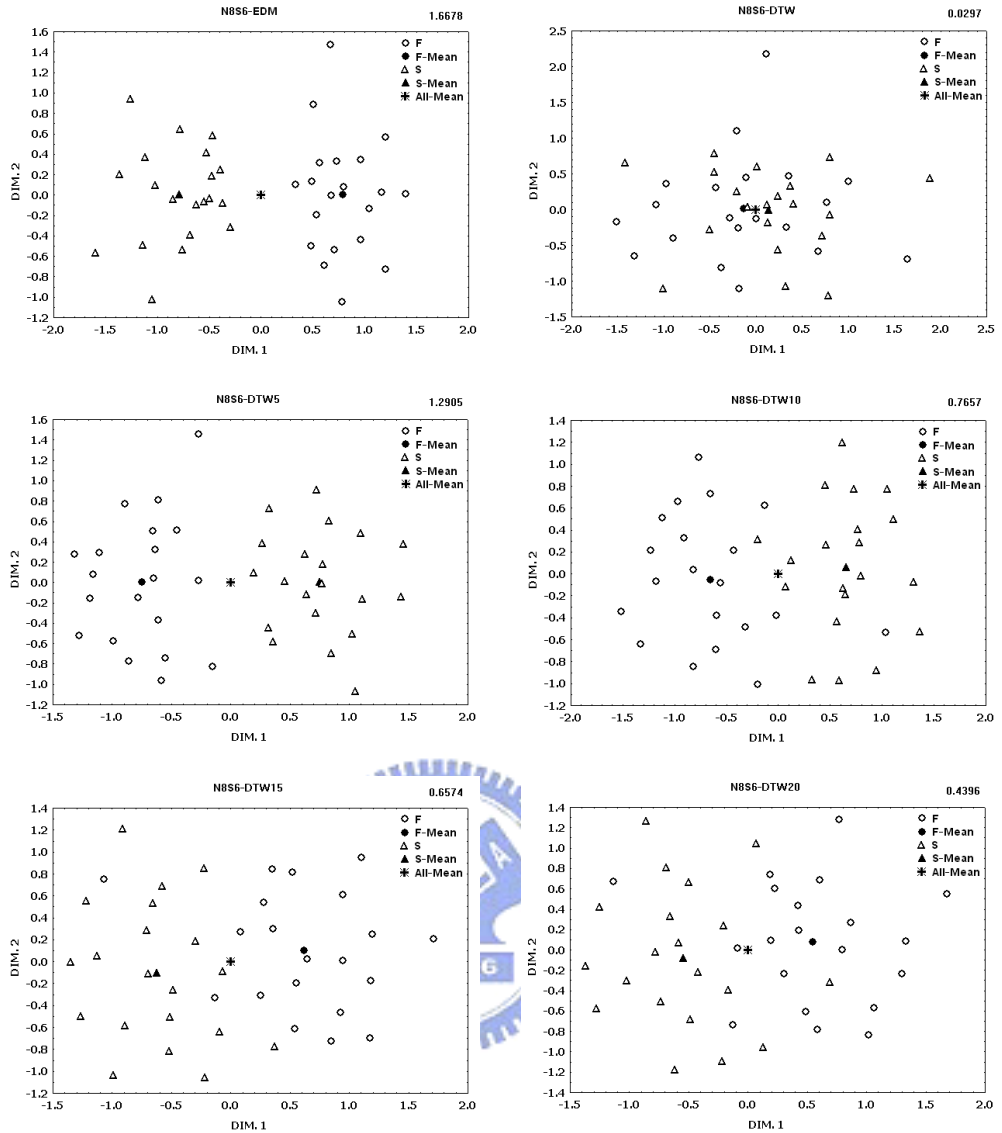


圖 4.13: N8S6之知覺圖

接下來探討本實驗的另一控制變數 — 雜訊大小，其對系統分類成效的影響。圖 4.12 為實驗所使用的大雜訊時間序列，雜訊在 ± 8 之內，同樣將資料分為固定與移動兩組。圖 4.13 則為“雜訊為 ± 8 以內，位移量為 6”的各相似度量測方法之知覺圖。

類別內大雜訊實驗的 F -試驗值數據列在表 4.2 中，圖 4.14 則展示了各相似度量測方法的 F -試驗值趨勢。同樣的，各方法的優劣度與小雜訊實驗所得結果相同，DTW5 曲線亦在位移量超過 5 時即急速上升。但比較兩實驗的數據可發現，大雜

表 4.2: N8與不同位移量下的各系統 F -試驗值

位移量	EDM	DTW	DTW5	DTW10	DTW15	DTW20
0	0.0363	0.0184	0.0039	0.0721	0.0844	0.0628
2	0.4284	0.0209	0.3008	0.3355	0.2463	0.2177
4	0.8822	0.0294	0.7579	0.5783	0.4139	0.3413
6	1.6678	0.0297	1.2905	0.7657	0.6574	0.4396
8	5.2536	0.0772	2.2932	1.2312	0.8331	0.5116
10	12.1310	0.0900	8.1404	2.2148	1.3873	1.0642

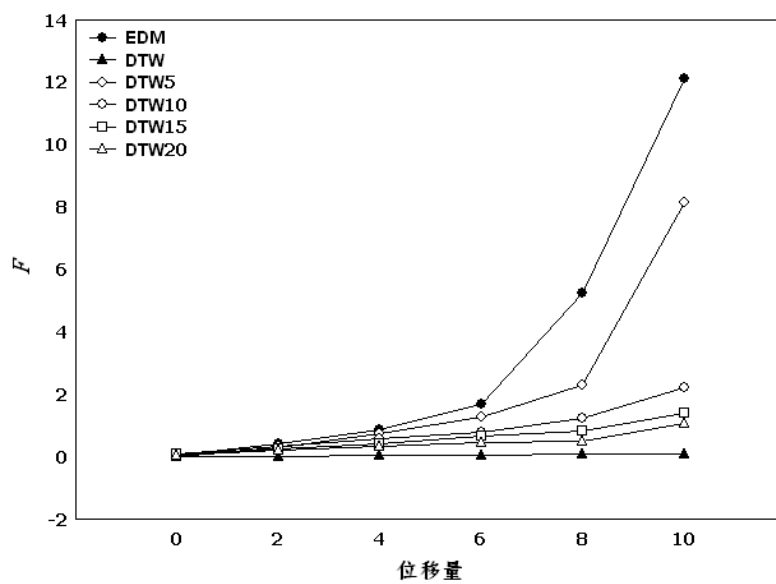


圖 4.14: N8與不同位移量下的各系統成效

訊實驗所得的 F -試驗值較小雜訊實驗來的小，在高位移量時，數據差距尤其顯著。這是因為資料樣板加上大雜訊後，所產生的時間序列之振幅很大，即使有位移情況發生，看起來差異仍不明顯，因此在分類時就不易將兩組資料分為不同的兩個類別。換句話說，若時間序列的走勢越混亂（上下抖動明顯），越不容易判讀其整體趨勢時，則在分類時就很容易跟它種類別資料混淆。

4.3 類別間實驗

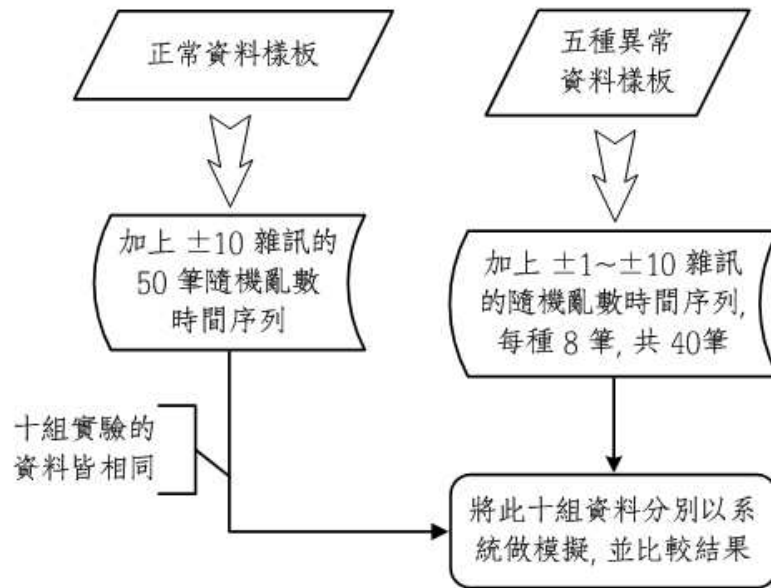


圖 4.15: 類別間實驗流程圖

類別間 (between categories) 實驗流程如圖 4.15 所示, 加在正常資料樣板上的雜訊在 ± 10 之間, 實驗所用之正常時間序列共 50 筆, 且此 50 筆資料將在之後的各組實驗數據中維持不變, 意即本類別間實驗只考慮更動異常類別資料。正常資料樣板與產生的時間序列, 以及正常狀態工作範圍和五種異常資料樣板, 參見節 4.1.1 的圖 4.2 至圖 4.4。前述步驟完成之後, 接下來就是以五種異常資料樣板為基底, 各別產生 8 筆異常時間序列, 如此即有 40 筆的五類異常時間序列。在產生異常時間序列時, 需分成十組數據, 即異常資料樣板上加上不同的雜訊大小 ($\pm 1 \sim \pm 10$), 如此再加上正常狀態的時間序列, 則每組數據皆有 90 筆資料。

本類別間實驗的目的是比較各相似度量測方法對此六種類別的分類成效, 包含在正常工作狀態下偵測異常時間序列, 以及將其判定 (分類) 為何種異常; 另外也探討雜訊大小對此實驗的影響, 因此本實驗的控制變數只有此兩種。不考慮位移量的影響, 是因為當錯誤類別位移時, 將更不相似於正常類別, 亦即偵錯將變的更容易, 如此顯而易見的结果不在本實驗的探討範圍內, 因此將類別內實驗的位移量控制因

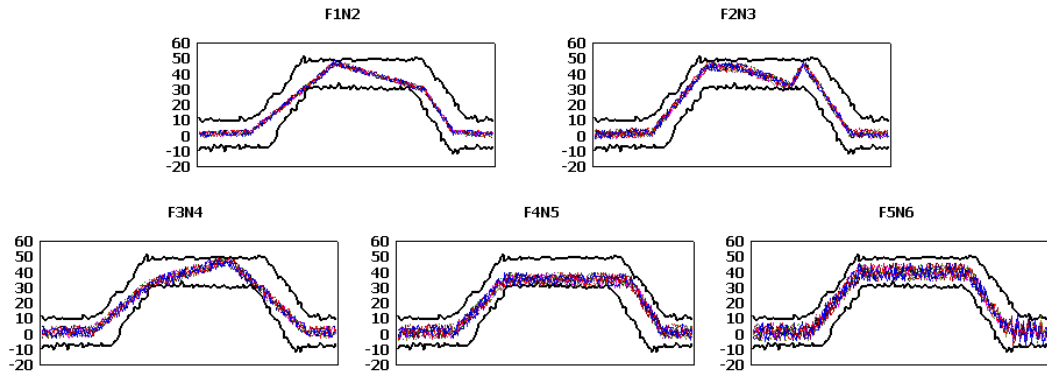


圖 4.16: 類別間實驗之異常時間序列

素去除。圖 4.16 展示部分的異常時間序列, 其中 F1N2 代表“第一類異常錯誤, 雜訊在 ± 2 之內”, 其餘依此類推。

圖 4.17 展示了雜訊在 ± 5 之內各相似度量測方法之實驗結果知覺圖, 圖中的 G 類別代表正常狀態類別, 而 F1 ~ F5 則分別代表五種異常狀態類別, 為使圖看起來較清楚明瞭, 並未將各類別與全體資料均值顯示在圖上。由圖中可得知, 當類別間分的越開, 類別內聚得越緊密, 則 F -試驗值就越大, 此與節 2.4 的推導結論相同。

由 EDM 的知覺圖可看出, 正常與異常狀態類別混淆在一起, 意即偵錯將較為困難。而另五種 DTW 方法的知覺圖中, 正常與異常狀態類別較能分辨出來, 因此使用 DTW 對於偵錯將有相當助益。此因使用 EDM 時, 由於大雜訊的關係 (參考節 4.2), 正常狀態類別與各異常狀態類別內的分佈太鬆散, 以至於所有的時間序列看起來都不太相似, 故很難將其分類。而使用 DTW 時, 正常狀態類別分佈的較緊密, 與異常狀態類別就較能區分開來。另外由圖可看出, 當 DTW 的視窗值越大時, 各異常類別內聚的較緊密, 且各異常類別間也較能區分, DTW5 ~ DTW20 的知覺圖尚不明顯 (此四圖的 F 值也較相近), 但 DTW 的圖中就可清楚看出各異常類別間的區分。

從 DTW 的知覺圖亦可發現, 異常狀態中與正常狀態最易區分的是第四類, 參考圖 4.4 可得知, 該類異常樣板的均值與正常樣板間偏離較多, 因此可容易的被偵

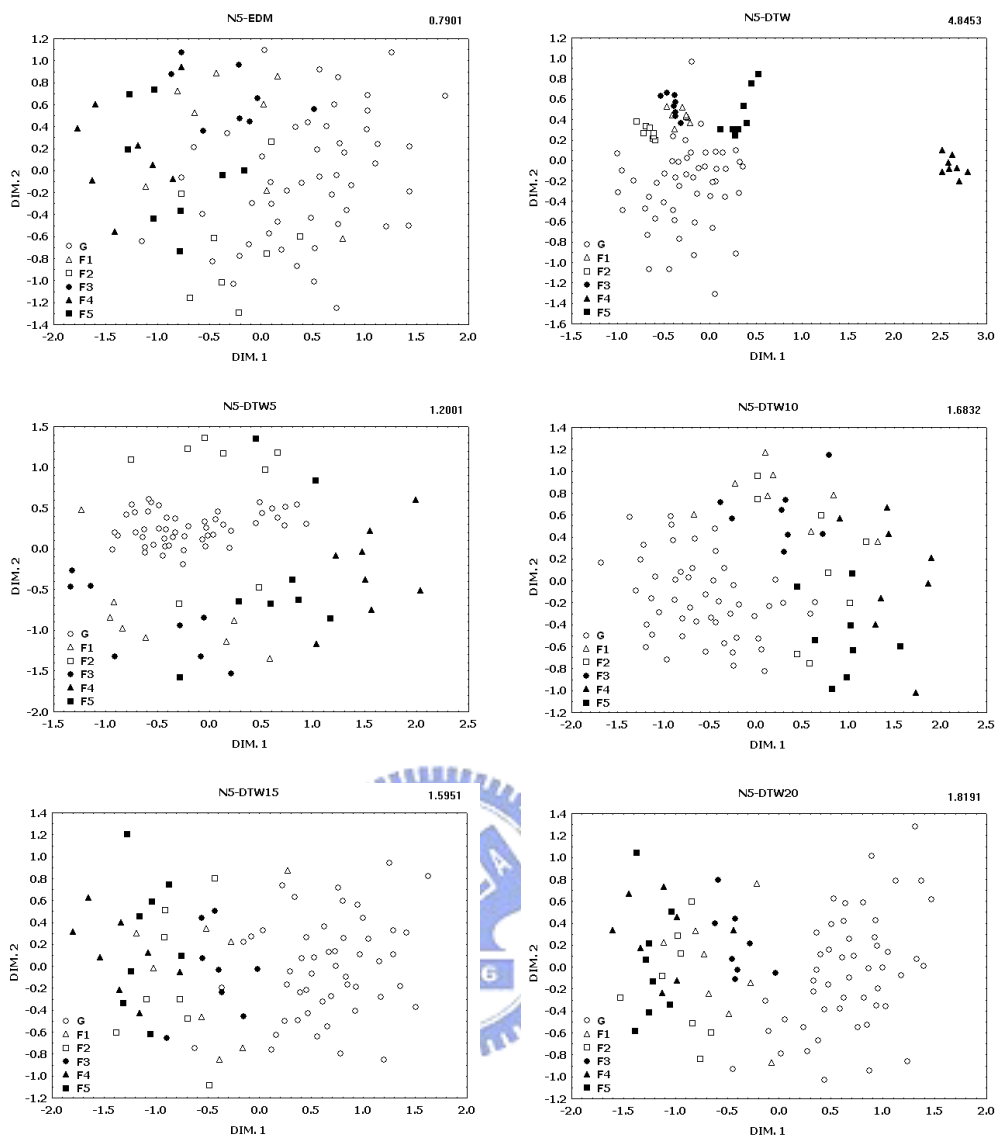


圖 4.17: N5之知覺圖

錯。而知覺圖也顯示第一類與第三類異常狀態較容易混淆，同樣由圖 4.4 可知，該兩類異常樣板的趨勢大致相同，都是中段突起的異常，差異處只是轉折點的不同，因此在使用 DTW 量測相似度時，因為兩類趨勢的相似而容易將其歸為一類。

表 4.3 列出了類別間實驗的 F -試驗值數據，圖 4.18 則展示各相似度量測方法在不同雜訊下的 F 值趨勢，為使圖形顯示與判讀方便，在此將 F -試驗值取對數。由表與圖可得知，隨著雜訊越大，各相似度量測方法的 F 值皆有下降的趨勢。這是因為當雜訊越大時，時間序列抖動的越厲害，使得各類別的差異相形之下變的較不顯

表 4.3: 不同雜訊大小下的各系統 F -試驗值

雜訊大小	EDM	DTW	DTW5	DTW10	DTW15	DTW20
±1	1.1560	1755.80	2.5589	4.0443	5.8031	7.9397
±2	1.1112	10.3260	1.9855	2.9334	4.1999	5.2545
±3	0.8074	6.0192	1.2373	1.7801	2.0053	2.3095
±4	0.6758	5.0298	1.2529	1.2800	1.8615	2.2501
±5	0.7901	4.8453	1.2001	1.6832	1.5951	1.8191
±6	0.5332	3.2470	0.9561	1.4168	1.7229	1.8804
±7	0.5347	3.7509	0.8239	0.9613	1.5697	2.0132
±8	0.3929	2.8002	0.7840	0.7601	1.2035	1.4044
±9	0.3798	1.7928	0.6204	0.7774	1.4115	1.4592
±10	0.1782	1.2740	0.2055	0.5003	1.0768	1.2631

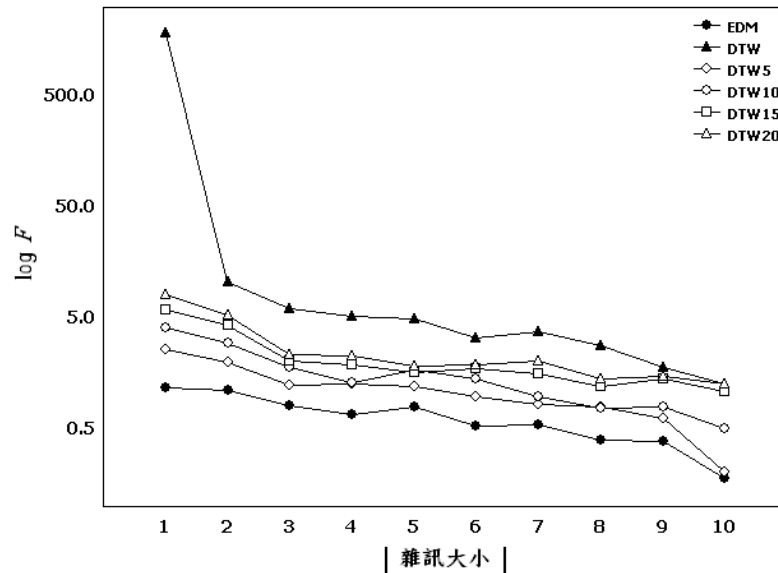


圖 4.18: 不同雜訊大小下的各系統成效

著, 因此各類別均值與全體均值近似 (S_B^2 變小), 意即類別間較難區分, 而類別內又分佈鬆散 (S_W^2 變大), 故使得 F -試驗值越來越小。類別間實驗的預期目標, 是希望 F -試驗值越大越好, 且對雜訊大小的影響較不敏感, 亦即 F 值曲線較平穩, 以此判定各相似度量測方法的系統成效可得: $DTW > DTW20 > DTW15 > DTW10 > DTW5 > EDM$, 與節 4.2 所得結論相同。

4.4 驗證分類結果

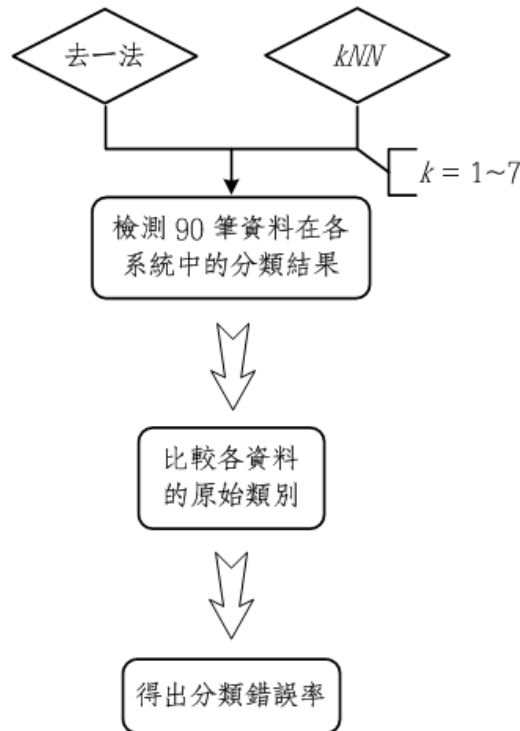


圖 4.19: 驗證流程圖

本論文將僅針對類別間實驗結果進行驗證，因類別內實驗的主要目的並不是判別分類結果錯誤與否，而是判別何種方法能有效量測位移後的時間序列仍屬於同一類別。驗證流程如圖 4.19 所示，“去一法”意指逐一從模擬資料集內取出時間序列，

表 4.4: N2之各系統分類錯誤率

k 值	EDM	DTW	DTW5	DTW10	DTW15	DTW20
1	0.256	0.000	0.078	0.067	0.022	0.011
2	0.256	0.000	0.078	0.067	0.022	0.011
3	0.278	0.000	0.122	0.056	0.022	0.022
4	0.278	0.000	0.111	0.044	0.011	0.011
5	0.300	0.000	0.100	0.022	0.011	0.011
6	0.289	0.000	0.100	0.033	0.011	0.011
7	0.322	0.000	0.078	0.033	0.000	0.000

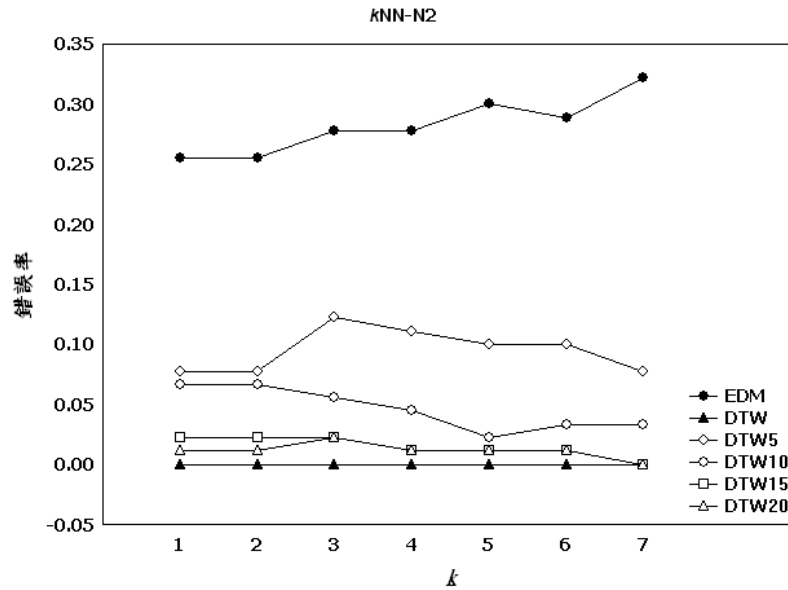


圖 4.20: N2之各系統分類成效

再利用加權式 k NN 判定該時間序列是屬於六種類別中的何類。由於去一法所用的時間序列是從原始資料集取出，因此我們已知該筆時間序列原是屬於何種類別，再對照系統的相似度量測與 k NN 的分類結果，將所有 90 筆資料驗證完後即可得知該系統模型的分類錯誤率。

表 4.4 與 4.5 分別列出當資料雜訊為 ± 2 與 ± 8 的分類錯誤率，而圖 4.20 與圖 4.21 各為此兩種雜訊時的各系統分類成效。當雜訊較小時， k 值對錯誤率的影響

表 4.5: N8之各系統分類錯誤率

k 值	EDM	DTW	DTW5	DTW10	DTW15	DTW20
1	0.433	0.289	0.344	0.300	0.211	0.256
2	0.433	0.289	0.344	0.300	0.211	0.256
3	0.433	0.311	0.367	0.289	0.311	0.289
4	0.433	0.300	0.400	0.300	0.289	0.300
5	0.444	0.322	0.467	0.322	0.333	0.300
6	0.444	0.300	0.456	0.322	0.333	0.322
7	0.444	0.322	0.478	0.367	0.367	0.322

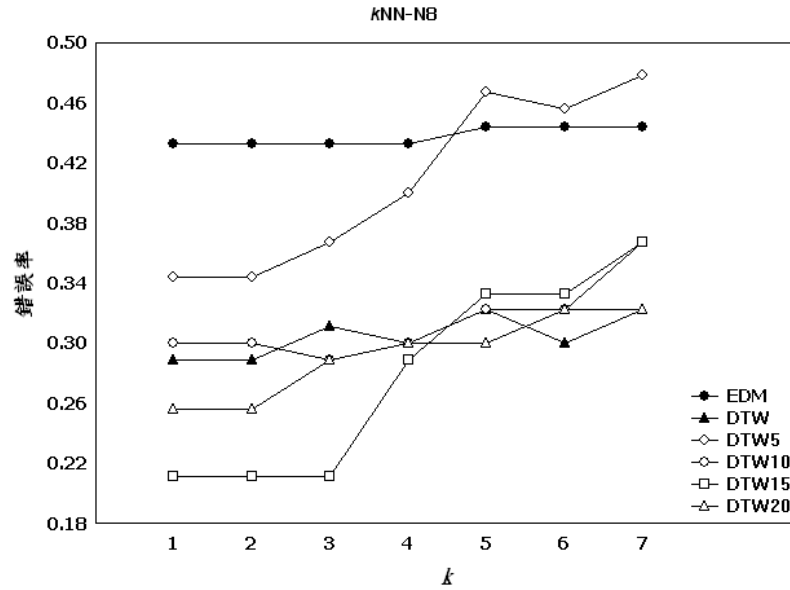


圖 4.21: N8之各系統分類成效

並不顯著，因為此時的系統分類成效較好（參考表 4.3），亦即各類別間較能區隔，且各類別內分佈較緊密，因此各方法的錯誤率曲線皆平穩。相較之下，當雜訊較大時，錯誤率會隨著 k 值增加而提高。此因雜訊較大時，系統分類的各類別間區隔不夠明顯，因此當 k 值過大時，就很容易將其他類別包含入分類範圍，進而造成分類錯誤。故可得一結論：當雜訊較小時， k 值的影響較不顯著；而當雜訊較大時， k 值將顯著影響分類成效，以圖 4.21 為例， $k \leq 3$ 較為恰當。另外不管雜訊大小，可發現 EDM 的分類錯誤率皆較 DTW 高出許多，依此驗證結果亦可得知 DTW 在分類時間序列上亦較 EDM 為佳。

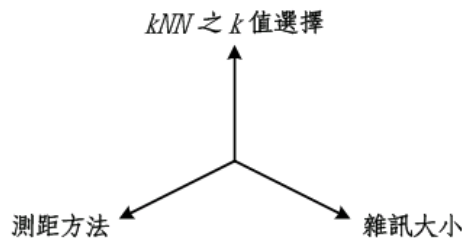


圖 4.22: 驗證之控制變數

整個驗證結果可由圖 4.22 的三種控制變數加以說明。表 4.6 至表 4.8 為 $k = 3$ 時的分類錯誤個數統計，當雜訊固定時，可看出 DTW 的分類錯誤率較 EDM 為低；而當相似度量測方法固定時，雜訊越大則錯誤率越高。以表 4.6 與表 4.8 來做說明，當雜訊小時，較容易分類錯誤的是正常類別，代表偵錯與分類異常訊號這兩部分主要目標有達成。所謂的雜訊大小是指加在異常類別樣板上的雜訊大小，而正常資料樣板所加的雜訊皆是固定在 ± 10 之間，因此相對於異常時間序列的小雜訊來說，正常時間序列就較容易錯誤分類。當異常時間序列的雜訊變大時，可發現正常類別皆分類正確，反而是異常時間序列分類錯誤，但表 4.8 只能說明異常類別的分類成效不佳，至於偵錯成效在此表無法說明，將在稍後的數據列表另行討論。

表 4.6: N1k3之各系統分類錯誤統計

N1k3	EDM	DTW	DTW5	DTW10	DTW15	DTW20
G	22	0	0	2	1	1
F1	0	0	0	0	0	0
F2	0	0	0	0	0	0
F3	0	0	0	0	0	0
F4	0	0	0	0	0	0
F5	0	0	3	0	0	0
小計	22	0	3	2	1	1

表 4.7: N5k3之各系統分類錯誤統計

N5k3	EDM	DTW	DTW5	DTW10	DTW15	DTW20
G	2	5	1	13	2	0
F1	7	0	7	2	0	0
F2	5	0	5	5	0	0
F3	6	0	4	0	0	5
F4	2	0	1	0	0	0
F5	6	0	7	8	7	1
小計	28	5	25	28	9	6

表 4.8: N9k3之各系統分類錯誤統計

N9k3	EDM	DTW	DTW5	DTW10	DTW15	DTW20
G	0	0	0	0	0	0
F1	8	5	8	8	4	6
F2	8	6	8	6	5	5
F3	8	2	8	6	8	7
F4	8	5	8	2	4	2
F5	8	7	7	7	6	6
小計	40	25	39	29	27	26

表 4.9: EDM 系統之分類錯誤統計

EDM	N1k1	N5k1	N9k1	N1k3	N5k3	N9k3	N1k5	N5k5	N9k5
G	20	3	0	22	2	0	19	2	0
F1	0	6	8	0	7	8	0	7	8
F2	0	4	8	0	5	8	0	6	8
F3	0	5	8	0	6	8	0	8	8
F4	0	2	8	0	2	8	0	3	8
F5	0	5	8	0	6	8	0	8	8
小計	20	25	40	22	28	40	19	34	40

表 4.10: DTW 系統之分類錯誤統計

EDM	N1k1	N5k1	N9k1	N1k3	N5k3	N9k3	N1k5	N5k5	N9k5
G	0	7	0	0	5	0	0	4	0
F1	0	0	5	0	0	5	0	0	5
F2	0	0	6	0	0	6	0	0	7
F3	0	0	1	0	0	2	0	0	3
F4	0	0	3	0	0	5	0	0	5
F5	0	0	7	0	0	7	0	0	7
小計	0	7	22	0	5	25	0	4	27

表 4.9 與表 4.10 分別列出 EDM 與 DTW 的分類錯誤個數情況，由表中數據可知，當相似度量測方式與 k 值皆固定時，錯誤率會隨著雜訊變大而增加，此與先

前所得結論相同。而當相似度量測方式與雜訊大小固定時，錯誤率有隨著 k 值變大而增加的趨勢。

表 4.11 至表 4.14 為幾個不同狀況與系統下的詳細分類數據，其中橫列代表每種類別時間序列的分類狀況，直行代表分類後各類別的時間序列個數。異常偵測錯誤率代表異常被歸類為正常的機率，整體分類錯誤率為所有樣本被分類錯誤的機率。表 4.11 與表 4.12 顯示，當異常類別的雜訊很小時，EDM 與 DTW 方法皆能有效偵測與分類異常時間序列，但因正常時間序列的雜訊較大，以至於 EDM 分類正常類別時會有較多錯誤出現，而此情況在 DTW 中沒有發生。再比較表 4.13 與

表 4.11: N1k3-EDM 之詳細分類表

N1k3-EDM	G	F1	F2	F3	F4	F5	正確	錯誤
G	28	1	0	0	0	21	28	22
F1	0	8	0	0	0	0	8	0
F2	0	0	8	0	0	0	8	0
F3	0	0	0	8	0	0	8	0
F4	0	0	0	0	8	0	8	0
F5	0	0	0	0	0	8	8	0
小計	28	9	8	8	8	29	68	22

異常偵測錯誤率: 0% 整體分類錯誤率:24.4%

表 4.12: N1k3-DTW 之詳細分類表

N1k3-EDM	G	F1	F2	F3	F4	F5	正確	錯誤
G	50	0	0	0	0	0	50	0
F1	0	8	0	0	0	0	8	0
F2	0	0	8	0	0	0	8	0
F3	0	0	0	8	0	0	8	0
F4	0	0	0	0	8	0	8	0
F5	0	0	0	0	0	8	8	0
小計	50	8	8	8	8	0	90	0

異常偵測錯誤率: 0% 整體分類錯誤率:0%

表 4.14 可發現, 當雜訊變大時, EDM 不但無法分類異常時間序列, 甚至還將其歸類為正常狀態, 亦即 EDM 的偵錯功能隨著雜訊變大而消失。反之, 雖然 DTW 的異常時間序列分類也不盡理想, 但其偵錯成效依然能維持, 以表 4.14 為例, 40 筆異常時間序列只有 3 筆被歸類為正常狀態, 偵錯成功率高達 92.5%。

表 4.13: N9k3-EDM 之詳細分類表

N1k3-EDM	G	F1	F2	F3	F4	F5	正確	錯誤
G	50	0	0	0	0	0	50	0
F1	8	0	0	0	0	0	0	8
F2	8	0	0	0	0	0	0	8
F3	8	0	0	0	0	0	0	8
F4	8	0	0	0	0	0	0	8
F5	8	0	0	0	0	0	0	8
小計	90	0	0	0	0	0	50	40

異常偵測錯誤率: 100% 整體分類錯誤率:44.4%

表 4.14: N9k3-DTW 之詳細分類表

N1k3-EDM	G	F1	F2	F3	F4	F5	正確	錯誤
G	50	0	0	0	0	0	50	0
F1	1	3	1	3	0	0	3	5
F2	2	1	2	2	1	0	2	6
F3	0	1	1	6	0	0	6	2
F4	0	2	0	0	3	3	3	5
F5	0	0	3	0	4	1	1	7
小計	53	7	7	11	8	4	65	25

異常偵測錯誤率: 7.5% 整體分類錯誤率:27.8%

第 5 章

研究結論

本論文探討如何在正常狀態的工作範圍條件下，偵測出異常的時間序列資料型態，並進一步將其分類，以期能對該異常做最迅速且正確的處理。本研究使用歐基里德距離量測法與動態時間扭曲法，直接比對時間序列相似度，並整合多重維度度量法與 k -最鄰近點分類法於本論文所發展的系統，並定義出一 F -試驗值函數作為系統成效的評估參考。

本論文以半導體製程中的質流控制器與氫氣為例，藉由其典型的正常與異常狀態作為資料樣板，自行模擬出實驗所需的時間序列資料集。相較於文獻既有與業界現行的方法，例如“工作範圍偵錯法”只能偵測出超出工作範圍的異常訊號，且無法對異常加以分類；而主成份分析法則因維度縮減的因素，常會損失資料的原有資訊，造成時間序列間微小的異常差異不易被偵測出來，進而導致分類的成效不佳；即使在正常狀態之工作範圍內的異常訊號，本系統仍能將之偵測出，且進而予以分類，這對業界製程上的變異偵測以及產品良率的提升上皆有莫大助益。本研究所發展出的系統除了應用在論文所舉的半導體工業之外，也可應用於資料型態與時間序列有關的議題，例如金融與醫學等領域。

對於時間序列相似度的量測，其中一個很重要的議題就是如何將“位移”這個變因的影響減至最低，意即尋找一套能判別整體時間序列趨勢的量測方法，在本研究中佔有舉足輕重的地位。從節 4.2 中的類別內實驗結果可清楚得知，動態時間扭曲法在減低位移因素影響這方面，的確較計算簡單的歐基里德距離量測法有效。而單就動態時間扭曲法來說，實驗結果也說明了該法的重要參數——“扭曲視窗長度”

對系統成效的影響，當設定的視窗長度越大，所能判別的時間序列趨勢範圍就更廣；換句話說，對位移因素的敏感度將越小，系統成效也就越好。

除了位移因素的考量外，任何業界實際資料都會碰到的問題，就是“雜訊”的影響。雖然本論文的时间序列不論異常與否，皆設定在正常狀態工作範圍之內，但雜訊大小仍對系統偵錯與異常分類之成效有重大影響。類別內實驗已可驗證使用動態時間扭曲法的系統，在降低位移因素對系統成效影響的這部份有良好表現，接下來就需驗證該法是否亦能降低系統對訊雜比 (Signal-to-Noise Ratio, SNR) 的敏感性。類別內與類別間實驗皆就這部份進行探討，從實驗結果可知，不論何種相似度量測方式，當異常類別的雜訊越大時，系統的異常偵測與分類成效皆會隨之變差。但不論訊雜比多大，動態時間扭曲法在降低系統的雜訊敏感度方面仍較歐基里德距離量測法優異，特別是當訊雜比較小時，動態時間扭曲法在偵測異常甚至於異常分類都有接近 100% 的正確率。而當訊雜很大時，本論文所使用的兩種主要量測方法在分類這部份的成效皆不好 (以表 4.13 與表 4.14 的 ± 9 雜訊之實驗數據為例，動態時間扭曲法的分類錯誤率接近 30%，而歐基里德距離量測法的分類錯誤率更接近 50% 左右。)，但動態時間扭曲法在偵測異常方面仍有超過 90% 正確率的表現。綜合上述結論可得，動態時間扭曲法除了在高訊雜比時的分類成效仍需改進之外，整體上在判別時間序列相似度方面仍為一良好方法。

最後，本論文有若干改進與發展的空間。首先，在時間序列資料集方面，由於未能取得業界半導體製程的實際資料，因此由虛擬資料所構成的本系統，在實際情況下能否有良好成效尚須進一步的驗證。其次，由於本實驗的多重維度量法藉助了 STATISTICA 統計軟體¹的運算，而其在讀入的矩陣資料上有最大 90×90 的尺寸限制，也就是說資料集最多只能有 90 筆時間序列，但為了正常與異常資料的比例 (正常資料稍多一些，以供正常工作範圍的擴展)，以及各異常類別的資料個數需相同等因素，使得本論文的每類異常資料只有 8 筆。樣本數上的不足很可能使系統模型建立時不夠準確，進而導致分類成效的錯誤率提升，因此若能增加時間序列的樣本數量，或可使系統對位移以及雜訊的敏感度更為減低。第三，本論文雖已使用

¹本研究使用第七版的 STATISTICA 統計軟體。

加權式 k -最鄰近點分類法來判別時間序列該屬於何類別, 但若能有比之更具分類正確性的方法輔助, 則系統的整體成效將更為顯著。第四, 本論文所做的是單變量研究, 如能將之擴展為多變量系統, 則在實際應用上將更為有效率。另外, 本論文的系統在每加入一筆新的時間序列作判讀時, 多重維度度量法所得的二維知覺圖會重新繪製, 亦即雖然系統模型中的樣本資料不變, 但不見得每次在比對新的時間序列時, 既有的樣本在知覺圖中的座標值亦會保持不變。雖然這部份並不影響結果 (因仍有 F -試驗值供成效評估, 知覺圖在本論文中主要作為一輔助評估工具), 但如能將系統建立時所得的二維知覺圖樣本座標值固定, 每當新判讀一時間序列時, 直接打點在既有的知覺圖上, 則在圖形的直覺判讀與系統成效評估上將更為方便。



參考文獻

- [1] S. Cunningham, C. Spanos and K. Voros, *Semiconductor Yield Improvement: Results and Best Practices*, IEEE Transactions on Semiconductor Manufacturing, Vol. 8, No. 2, 1995.
- [2] S. Benninga and B. Czazkes, *Financial Modeling*, MIT Press, Massachusetts, 1997.
- [3] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*, Springer, New York, 2001.
- [4] *Workshop on the Application of Advanced Process and Equipment Control in Semiconductor Manufacturing*, Department of Industrial Technology, Ministry of Economic Affairs, Taipei, 2003.
- [5] H. Xiao, *Introduction to Semiconductor Manufacturing Technology*, Prentice Hall, New Jersey, 2000.
- [6] P. V. Zant, *Microchip Fabrication: A Practical Guide to Semiconductor Processing*, 4th ed., McGraw-Hill Professional Publishing, Columbus, 2000.
- [7] G. W. Rubloff, *Challenges for Real-Time Control in Reactive Semiconductor Manufacturing Process Environments*, http://www.isr.umd.edu/gwrubloff/research/recent_results/

presentations/presentations-pdfs/AVS98.Invited.talk.pdf#
search=%22%22MFC%20shutdown%22%22

- [8] S. Raychaudhuri, J. M. Stuart and R. B. Altman, *Principal Components Analysis to Summarize Microarray Experiments: Application to Sporulation Time Series*, Pacific Symposium on Biocomputing, Vol. 5, Pages 452–463, 2000.
- [9] N. B. Gallagher and B. M. Wise, *Development and Benchmarking of Multivariate Statistical Process Control Tools for a Semiconductor Etch Process: Improving Robustness through Model Updating*, Control Engineering Practice, Vol. 12, Pages 857–862, 2004.
- [10] F. Jia, E. B. Martin and A. J. Morris, *Non-Linear Principal Components Analysis with Application to Process Fault Detection*, International Journal of Systems Science, Vol. 31, No. 11, Pages 1473–1487, November 2000.
- [11] N. F. Thornhill, S. L. Shah and B. Huang, *Detection of Distributed Oscillations and Root-Cause Diagnosis*, Proceedings of the CHEMFAS4, Pages 167–172, July 2001.
- [12] A. Basilevsky, *Statistical Factor Analysis and Related Methods, Theory and Applications*, John Wiley & Sons, New York, 1994.
- [13] B. Everitt and G. Dunn, *Applied Multivariate Data Analysis*, Oxford University Press, New York, 1992.
- [14] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, Pages 92–110, 1986.
- [15] G. Strang, *Introduction to Linear Algebra*, 3rd ed., Wellesley-Cambridge Press, Massachusetts, 2003.

- [16] H. F. Kaiser, *The application of electronic computers to factor analysis*, Education and Psychological Measurement, Vol. 20, Pages 141–151, 1960.
- [17] T. W. Anderson, *The Statistical Analysis of Time Series*, John Wiley & Sons, New York, 1971.
- [18] R. Cattell, *The Scree Test for the Number of Factors*, Multiv. Behav. Res., Vol. 1, 1966.
- [19] J. Michell, *An Introduction to the Logic of Psychological Measurement*, Lawrence-Erlbaum Associates, New Jersey, 1990.
- [20] *Scaling Method*, <http://www.psy.ntu.edu.tw/pm/method/Scaling.htm>
- [21] G. Churchill and D. Iacobucci, *Marketing Research: Methodological Foundations*, 9th ed., South-Western College Pub., Cincinnati, 2004.
- [22] S. Schiffman, M. Reynolds and F. Young, *Introduction to Multidimensional Scaling: Theory, Methods and Applications*, Academic Press, 1981.
- [23] I. Borg and J. Lingoes, *Multidimensional Similarity Structure Analysis*, Springer, New York, 1987.
- [24] *Torgerson's Classical MDS Derivation*, <http://forrest.psych.unc.edu/teaching/p230/Torgerson.pdf>
- [25] J. Costa, N. Patwari, and A. Hero III, *Distributed Weighted-Multidimensional Scaling for Node Localization in Sensor Networks*, ACM Transactions on Sensor Networks, Vol. 2, No. 1, Pages 39–64, February 2006.
- [26] J. Lattin, J. D. Carroll and P. E. Green, *Analyzing Multivariate Data*, Thomson Learning, Missouri, 2003.

- [27] *Instance-Based Learning*, http://www.it.nuigalway.ie/m_maddent/ct433/pdf/ML_03_IBL.pdf
- [28] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Kluwer-Academic Publishers, London, 1998.
- [29] M. Christophe, *Neural Networks for Pattern Recognition*, Oxford, UK, 1995.
- [30] Finding Similar Time Series, <http://appsrv.cse.cuhk.edu.hk/~mzhou/Classified%20References/Survey/Time%20series%20similarity%20measures.ppt#525,1,FindingSimilarTimeSeries>
- [31] D. J. Berndt and J. Clifford, *Using Dynamic Time Warping to Find Patterns in Time Series*, In Working Notes of the Knowledge Discovery in Databases Workshop, Pages 359–370, July 1994.
- [32] E. J. Keogh and M. J. Pazzani, *A Simple Dimensionality Reduction Technique for Fast Similarity in Large Time Series Databases*, Knowledge Discovery and Data Mining, Current Issues and New Applications, 4th Pacific-Asia Conference, Pages 122–133, 2000.
- [33] T. S. F. Wong and M. H. Wong, *Efficient Subsequence Matching for Sequences Databases under Time Warping*, Database Engineering and Applications Symposium, 2003. Proceedings, 7th International, Pages 139–148, July 2003.
- [34] E. J. Keogh and M. J. Pazzani, *Scaling up Dynamic Time Warping to Massive Datasets*, 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases, Pages 1–11, 1999.
- [35] E. J. Keogh and M. J. Pazzani, *Scaling up Dynamic Time Warping for Datamining Applications*, Conference on Knowledge Discovery in

Data, Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Pages 285–289, August 2000.

[36] 工業技術研究院 機械工業研究所, IC 製造前段設備關鍵零組件專題研究 — 真空邦浦、質流控制器, 1997。

[37] 黃俊英, 多變量分析, 中國經濟企業研究, 台北, 2000。

[38] 文件內容自動分類, <http://www.lins.fju.edu.tw/~tseng/ResearchResults/categorization.htm>

