

行政院國家科學委員會補助產學合作研究計畫成果完整報告

應用資料採礦於台灣中小尺寸面板廠品質異常原因分析模式

計畫類別： 先導型 開發型 技術及知識應用型

計畫編號：NSC 99-2622-E-029-007-CC3

執行期間：99年06月01日至100年05月31日

執行單位：東海大學工業工程與經營資訊學系

計畫主持人：洪堯勳

共同主持人：

計畫參與人員：

處理方式：依規定，原則上不予公開。

中 華 民 國 100 年 08 月 29 日

應用資料採礦於台灣中小尺寸面板廠品質異常原因分析模式

摘要

液晶顯示器面板製造程序相當繁複且製造速度快速，若是未能掌握品質異常之原因並及時改進，則會造成大量不良品的產生，不利成本之競爭。

本研究針對台灣某家中小尺寸液晶顯示器面板製造廠，利用資料採礦技術進行品質問題歸因方面之探討，首先使用二階段分群法及自我組織映射圖網路法，以比較不同方法的正確率，其次利用正確率較高的結果，使用決策樹方法歸納出造成品質不良現象的法則，可提供廠方快速找出品質異常原因。

研究結果顯示，經過分群後的資料的確提高判斷上的正確率，並且利用自我組織映射圖網路得到之結果會比二階段分群法為佳。在決策樹分析方面，本研究證實可以經由此方法得到判斷品質問題的法則，廠商可利用相關手法不斷更新法則，以掌握最及時的品質狀況。

關鍵字詞：資料採礦、集群分析、決策樹、品質問題分析、液晶顯示器面板

Applying Data Mining to the Quality Analysis Model in Small and Medium Size TFT-LCD Industry

ABSTRACT

The manufacture process of TFT-LCD is very complicated and the manufacture speed is also fast. It will bring a large number of defective goods and make the cost up if we don't master the reasons of quality unusual and to reform them.

We try to use data mining to research the attribution of a Taiwan small and medium size TFT-LCD industry's quality problem. First, we use Two-Steps Cluster and Self-Organizing Maps and compare the correct rate between them. Second, based on the higher correct rate clustering method, we use decision tree to generalize the rules that cause the quality problem. According to these rules, industries can find out the reason of quality problem swiftly.

The research result shows that the correct rate will be higher after clustering and the result of SOM will be better than Two-Steps Cluster in clustering outcome. After that, we also verify that we can obtain the rules to determine the quality problems by using decision tree. Industries can use these methods to renovate rules so that the newest condition of quality can be commanded.

Keywords: Data Mining 、 Two-Step Cluster 、 SOM 、 Quality Analysis 、 TFT-LCD

目錄

摘要	II
ABSTRACT	III
誌謝	錯誤! 尚未定義書籤。
目錄	IV
圖目錄	VI
表目錄	VII
第一章 緒論	1
1.1 研究背景與動機	1
1.2 研究問題與目的	2
1.3 研究方法	2
1.4 研究範圍與限制	3
1.5 研究架構與流程	3
第二章 文獻探討	5
2.1 中小尺寸 TFT-LCD 之產業說明	5
2.2.1 中小尺寸 TFT-LCD 之產業背景	5
2.1.2 TFT-LCD 之產業特性	6
2.1.3 TFT-LCD 之相關研究	8
2.2 中小尺寸 TFT-LCD 製造程序與品質問題	9
2.2.1 TFT-LCD 製造流程	9
2.2.2 TFT-LCD 品質檢測技術	10
2.2.3 TFT-LCD 品質問題及相關研究	12
2.3 資料採礦	14
2.3.1 資料採礦之定義及程序	14
2.3.2 資料採礦相關技術及功能	16
2.3.3 資料採礦應用於 TFT-LCD 之相關研究	19
2.4 本章小結	21
第三章 研究方法	22
3.1 類神經網路	22
3.1.1 類神經網路之定義及原理	22
3.1.2 數學式	23
3.1.3 網路的運作方式	24
3.1.4 學習法則—監督式學習與無監督式學習	24
3.2 集群分析	24
3.2.1 自組織映射網路	25
3.2.2 二階段分群法	29
3.3 決策樹	33
第四章 實證分析	39
4.1 問題定義與資料屬性定義	39

4.1.1 問題定義.....	39
4.1.2 資料準備與屬性定義.....	39
4.2 決定分群數目.....	44
4.2.1 華德法初步分群.....	44
4.2.2 K 平均數法分群.....	46
4.2.3 對 K 平均數法結果做驗證.....	49
4.3 分群結果.....	52
4.3.1 K-means 分群結果.....	52
4.3.2 自我組織映射圖分群結果.....	55
4.4 決策樹分析.....	60
4.3 本章小結.....	66
第五章 結論與建議.....	67
5.1 結論.....	67
5.1.1 方法面結論.....	67
5.1.2 分析面結論.....	68
5.2 未來研究與建議.....	68
參考文獻.....	70

圖目錄

圖 1.1 研究流程.....	4
圖 2.1 中小尺寸面板產值與產量.....	6
圖 2.2 中小尺寸主要應用出貨量.....	6
圖 2.4 資料採礦流程.....	15
圖 2.5 資料採礦的重要步驟.....	16
圖 3.1 研究方法架構圖.....	22
圖 3.2 生物神經元.....	23
圖 3.3 人工神經元.....	23
圖 3.4 SOM 的空間映射概念.....	26
圖 3.5 SOM 架構.....	26
圖 3.6 優勝單元與鄰近區域內權重修正圖.....	29
圖 3.7 決策樹分支結構.....	33
圖 3.8 ID3 決策樹歸納結果示例.....	35
圖 4.1 造成線不良之特性要因圖.....	41
圖 4.2 K-means 疊代收斂圖.....	48

表目錄

表 2.1 TFT-LCD 在製造程序方面文獻.....	8
表 2.2 TFT-LCD 檢測設備分類.....	11
表 2.3 檢測工程的概要.....	12
表 2.4 TFT-LCD 有關品質研究相關文獻.....	14
表 2.5 分類與分群之說明與技術.....	18
表 2.6 資料採礦應用於 TFT-LCD 之相關研究.....	20
表 3.1 條件 X 屬於分類 C 的對應表.....	38
表 4.1 參數編碼示例.....	42
表 4.2 人員原因變數.....	42
表 4.3 機器原因變數.....	43
表 4.4 材料原因變數.....	43
表 4.5 方法原因變數.....	43
表 4.6 環境原因變數.....	44
表 4.7 群數凝聚過程.....	44
表 4.8 二集群因子平均數.....	45
表 4.9 五集群因子平均數.....	45
表 4.10 初始集群中心點.....	47
表 4.11 疊代記錄.....	47
表 4.12 最後集群中心點.....	48
表 4.13 各集群中的觀察值個數.....	49
表 4.14 分群結果 ANOVA 表.....	49
表 4.15 檢定結果 ^a	50
表 4.16 Wilks' Lambda 值.....	50
表 4.17 標準化的典型區別函數係數.....	51
表 4.18 各組重心的函數.....	52
表 4.19 K-means 第一群之品質類型及紀錄編號.....	52
表 4.20 K-means 第二群之品質類型及紀錄編號.....	53
表 4.21 K-means 第三群之品質類型及紀錄編號.....	54
表 4.22 K-means 第四群之品質類型及紀錄編號.....	54
表 4.23 K-means 第五群之品質類型及紀錄編號.....	54
表 4.24 SOM 的學習參數設定.....	55
表 4.25 SOM 第一群之品質類型及紀錄編號.....	56
表 4.26 SOM 第二群之品質類型及紀錄編號.....	57
表 4.27 SOM 第三群之品質類型及紀錄編號.....	58
表 4.28 SOM 第四群之品質類型及紀錄編號.....	59
表 4.29 SOM 第五群之品質類型及紀錄編號.....	60
表 4.30 第一群決策樹分類法則.....	61
表 4.31 第二群決策樹分類法則.....	62

表 4.32 第三群決策樹分類法則.....	63
表 4.33 第四群決策樹分類法則.....	64
表 4.34 第五群決策樹分類法則.....	65

第一章 緒論

1.1 研究背景與動機

近幾年來液晶顯示器(Liquid Crystal Display, LCD)蓬勃發展，其中又以薄膜電晶體液晶顯示器(Thin Film Transistor-Liquid Crystal Display, TFT-LCD)為最大宗，本研究所要探討的中小尺寸顯示面板(一般為 10.4"以下)應用的領域與大尺寸面板不同，在技術及品質上的要求也有所差異。在需求方面，小尺寸面板必須符合輕薄短小的要求。在生產方面，中小尺寸顯示面板市場屬於客製化市場(Customer design)，其面板尺寸規格多樣化。而在目前的 TFT-LCD 產業中，台灣與韓國因策略定位在「Process Technology」同屬於面板製造，因此台、韓廠商競爭最為激烈(蔣漢旗, 2002)。韓國因產業上下游一貫生產發展已久，關鍵零組件已經有極高比例國產化，不需花費外匯向海外進口，因此在同樣材料成本之下，韓國 LCD 面板的外銷競爭力大增，自 2005 開始已經與台灣並列世界一二名(王綺年, 2008)，並且可以預期的是中國大陸廠商加入 TFT-LCD 生產的競逐行列。在此情形之下，台灣廠商除了交貨快速的優點外(王綺年, 2008)，若是能從提升產品品質的工作著手，用更精良的產品來迎戰，將會是提高本身產品在國際上競爭力的做法之一。

TFT-LCD 產品的製造程序相當繁複，在這種具高度精密技術的生產環境下，掌控制程不良率（或良率）的變動是相當重要的工作。但是在產品品質不佳的情況下欲改善品質，卻缺乏一套全面瞭解品質異常發生原因的分析架構來做為改善的依循的問題。

TFT-LCD 的生產速度通常很快，若發現品質不良必須馬上找出造成品質不良的原因並加以改善，以免不良品被重複製造。然而造成不良的因素太多太雜，往往耗費許多時間才能找到問題根源，若是品質異常原因分析人員每一個可能成因都去檢查，可能造成發現異常原因速度過慢造成未能及時對不良原因加以改進，因此本研究主要希望提出法則讓分析人員能夠快速找出造成不良現象的原因，減少不良品被製造出來的數量，進而提升產品品質。

薄膜型液晶顯示器(TFT-LCD)共有前段陣列(array)製程、LCD面板組裝(cell)製程與模組組裝三個主要的製程，其中前段陣列製程與LCD面板組裝製程是薄膜型液晶顯示器(TFT-LCD)最重要的製程，因為在前段陣列與面板組裝製程中，如果品質發生差異，則會造成面板厚度不均勻、間隙物密度不均勻或是蝕刻線寬太寬或太窄，將使得在後段模組組裝或是產品功能上發生問題，因而導致成本的浪費，本研究希望以CELL製程作為研究對象，整理出能迅速辨認造成品質問題的快速途徑。

1.2 研究問題與目的

目前中小尺寸TFT-LCD若發生品質不良現象，品質異常分析人員必須要每一項可能成因都去檢測，導致檢測時間拉長，未能及時對造成品質不良現象的原因做更正，所以本研究的重點在於建立一個分析模式，對於不良現象的結果做一個回溯性的探究，找出影響品質不良現象之最關鍵的因素。本研究所提出的法則，在遇到相同或類似問題時可以立即且快速的找出品質問題根源，達成下列目標：

- 1.提出尋找品質不良原因的方法：任何產業都會有學習的現象發生(Hirschmann,1964)，本研究提出一個可以快速判斷是哪些因素導致不良現象產生的方法，重要的是此方法提出之法則並不能長久沿用，在學習效果下，品質必定會持續進步，而這種改變也導致了法則的增加或減少，因此本研究之法則必須隨著技術進步而不斷重複實驗並更新。
- 2.快速了解到品質異常發生之原因：本研究之個案公司可以透過此項研究來尋找造成品質不良的原因，不同於過去的分析方式，本研究所進行之分群及決策樹分析可以判斷最有可能造成品質問題發生的原因，直接去檢查該原因應可節省許多逐項檢查的時間。

1.3 研究方法

資料採礦(Data Mining)的定義最早由Fayyad於1991年所提出，其目的是從龐大的資料中找出規則。Han & Kamber(2001)提出資料採礦即是在資料庫中，利用各種分析方法與技術，將過去所累積的大量歷史資料，進行分析、歸納、預測與整合，尋找出有價值的隱藏事件，加以分析，以萃取出有用的資訊或找出使用者有興趣的樣式與知識，提供企業管理階層，作為訂定決策的依據。本研究即運用資料採礦中較常被使用做為找出規則之分

群及分類方法做為分析方法，並利用Balakrishnan et al.(1994)提出的觀點，針對正確率較高的分群方法得到之結果，利用決策樹(decision tree)的方法針對品質問題原因做探究，期望能得到最直接的結果，在品管人員判斷品質問題時，能第一時間就找到問題的根源，進而減少不必要之浪費。

1.4 研究範圍與限制

目前發展TFT-LCD產業的國家包含日本、韓國、台灣以及最近積極發展的中國。而本研究基於各地區產業發展的完整性以及資料的易取性，選擇以台灣地區之某一家中小尺寸TFT-LCD廠作為研究對象，透過訪談與資料蒐集，建立一個可供廠商在發生品質問題時參考的架構。

而在研究限制方面有下列幾點：

- 1.TFT-LCD屬於高科技產業，在技術上也是日新月異。在技術的發展過程中，可能會因為技術上的突破或進步，而對面板廠的品質不良原因產生影響。由於無法事先掌握TFT-LCD技術上的變化或趨勢。因此，本研究無法立即考慮技術上的變更，此為研究的限制之一。
- 2.本研究僅以一家TFT-LCD廠商為研究對象，其技術程度可能會影響其品質良莠，而此標準可能在其他廠商之製造過程中不適用，必須加以重新定義。
- 3.如製程變動規模過大可能會影響最後效果，因此本研究之個案廠商必須具有下列特點：生產穩定度高，沒有重大的突發事故；生產期間沒有大規模更動生產設備、作業流程與工具設計；使用之原物料品質穩定度高；作業人員皆受過訓練且流動率不高也不考慮生產中斷、生產遺忘與重製的學習效應。

1.5 研究架構與流程

本論文共分為五章，其架構內容分述如下：

第一章:說明本研究所根據之背景，提出研究動機與目的，針對研究方法做說明並提出研究所包含的範圍及所會遇到的限制。

第二章:針對中小尺寸 TFT-LCD 產業做現況說明，包括產業背景與產業特性，接著利用製造原理帶出品質相關問題並對 TFT-LCD 品質檢測過程做一說明，接著再對本研究所採用之研究方法資料採礦做簡略說明。

第三章:本章針對本研究採用的資料採礦中的分群方法「自組織映射圖網路」及「二階段分群法」做處理程序上的說明，最後對找出規則之用的「決策樹」方法做說明。

第四章:說明個案廠商在蒐集品質問題之程序，對蒐集來的變數做整理及編碼的動作，接著對所分出的各群做意義上的解釋，歸納品質問題發生原因，說明決策樹歸納法則之產生，初步建立品質問題診斷系統。

第五章:從實例分析之結果，提出結論、建議與未來研究重點與方向。

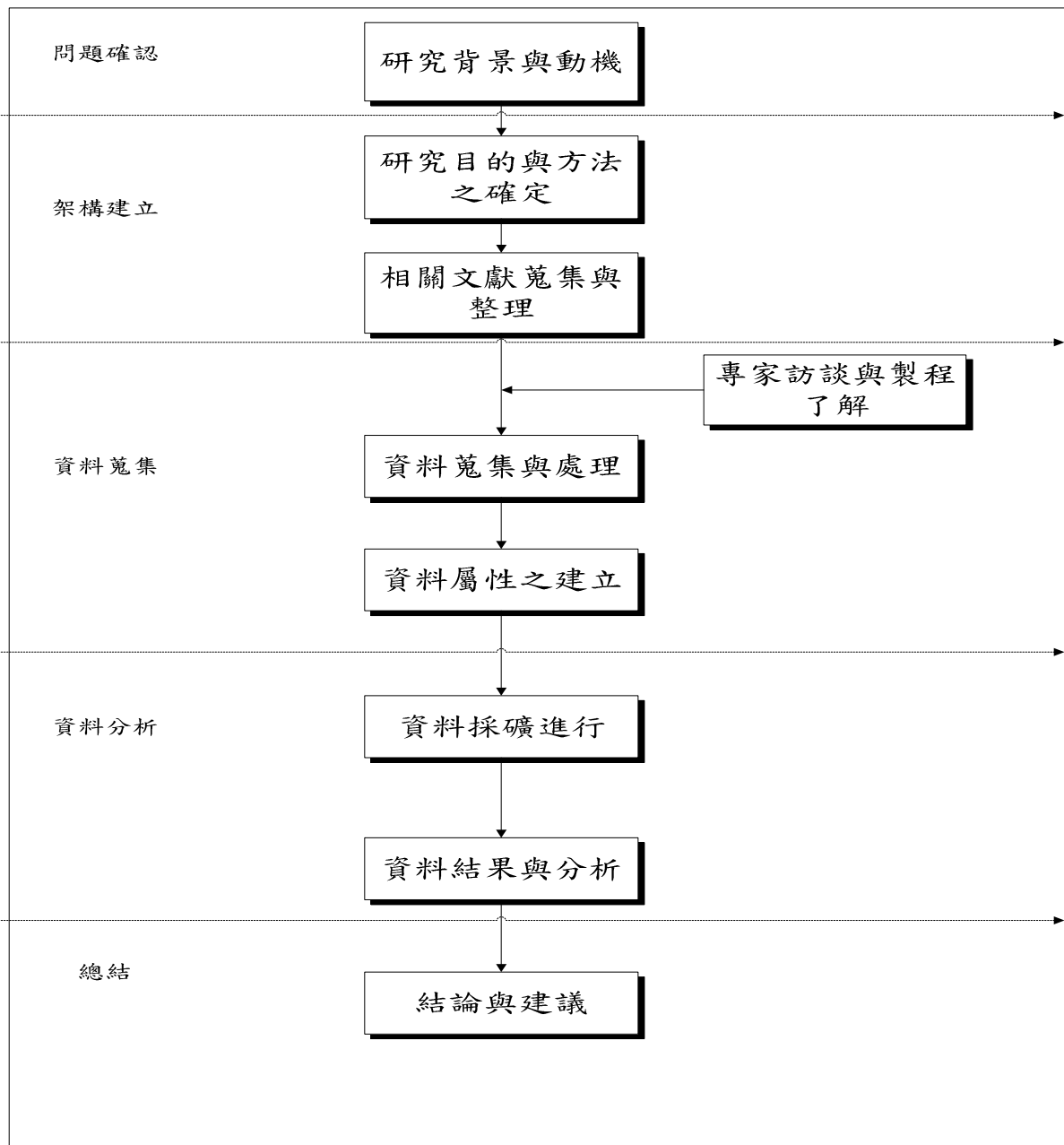


圖 1.1 研究流程
(資料來源:本研究整理)

第二章 文獻探討

本章首先簡略說明目前中小尺寸 TFT-LCD 之產業背景，並探討此產業的特性以彰顯品質的重要性。接著對製造原理及流程做約略說明，以及過去研究對此產業有關品質之相關研究。最後針對本研究所使用的方法資料採礦做介紹及流程探討，說明為何本研究要使用分群及分類兩種方法，探討過去有哪些研究也是使用類似的方法試圖歸納出造成品質問題的法則，並針對 TFT-LCD 產業做過去研究之蒐集與分析。

2.1 中小尺寸TFT-LCD 之產業說明

2.2.1 中小尺寸TFT-LCD 之產業背景

亞洲三國之 TFT- LCD 產業競爭力，其中我國以資金供應、筆記型電腦之內需市場、其他下游應用之內需市場、生產成本優勢、市場反應彈性為競爭優勢之來源。這些優勢皆源自國內 IC 與資訊產業發展所產生之附加價值，對於影響我國 TFT-LCD 產業之關鍵因素如量產經驗、產業結構完整性、科技人員之充裕度、產品開發部分皆相較於日韓缺乏，此亦為我國 TFT-LCD 產業發展亟待努力之部分。魏永康(2004) 採用個案研究法，選擇臺灣五大 TFT-LCD 面板廠商（友達、奇美、華映、瀚宇彩晶、廣輝）作為研究對象。結果發現，台灣廠商之主要競爭優勢在於彈性之製造管理能力及良好之市場行銷通路，積極進行垂直策略聯盟以尋求與供應商及銷售通路間的垂直鏈結，使得相關的價值活動達到最佳化。業者必須掌握企業核心競爭力、認清液晶循環趨勢及擬定企業競爭策略；進而尋求企業合併與上中下游產業的垂直整合。

而中小尺寸面板市場發展迥異於大尺寸面板，除一般認為其客制化程度高，技術門檻亦高外，各種不同面板技術間相互競爭，且又各具利基市場也是產業的一大特色。除 TFT-LCD 面板外，低階產品的彩色 STN 面板目前仍位居出貨量大宗；而高階部分則有被動式及主動式 OLED 面板；甚至 TFT-LCD 面板本身也還有 a-si 及 LTPS 等技術的區別。此外，各家廠商更是不斷開發獨有技術，推出與其他競爭者區隔的產品。由於 TFT-LCD 面板技術的不斷演進及價格下降，使得不論是 a-si 或 LTPS 技術在中小尺寸面板的滲透率不斷提昇。根據預估，2005 ~ 2010 年 TFT-LCD 中小尺寸面板銷售量

CAGR 16.05%、產值 CAGR 9.58% 均較整體中小尺寸面板成長率為高，顯示TFT-LCD面板滲透率將會向上持續攀昇。

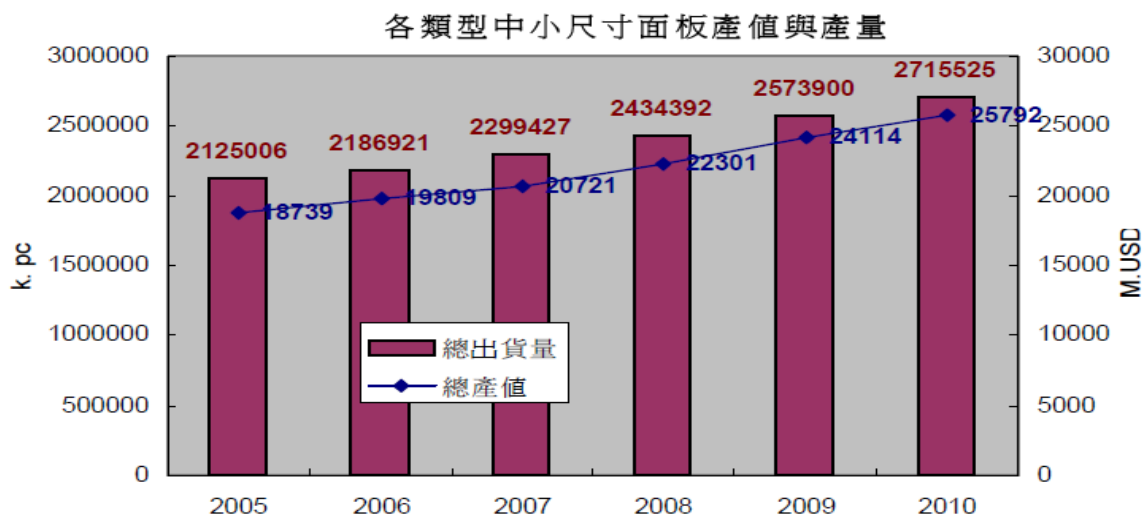


圖 2.1 中小尺寸面板產值與產量

(資料來源:台灣工業銀行(IBT) 2008/1)

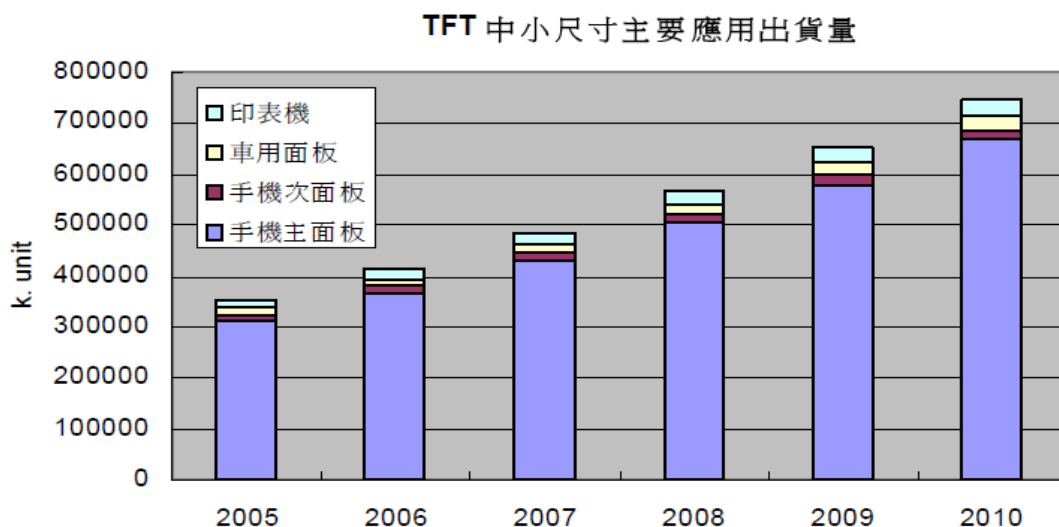


圖 2.2 中小尺寸主要應用出貨量

(資料來源:台灣工業銀行(IBT) 2008/1)

2.1.2 TFT-LCD 之產業特性

根據王昭琪（2002）整理，TFT-LCD 之產業特性有以下的特性：

一、資本密集，廠商進入障礙與退出成本高

以一個大型TFT-LCD廠商而言，其建廠資金從技術研發到量產階段約需投資至少百億元以上，且因固定資產之投入相當龐大，故欲達到經濟規

模之產量（ESQ）也相對提高，進入障礙也相對提昇。然而廠商一旦投入此產業之固定資產將成為沈沒成本，廠商若欲退出此產業除了出售事業體，將僅有承擔此沈沒成本一途，故過高的沉沒成本形成退出門檻。

二、技術密集—智財權（IPR）形成法律性的進入障礙

TFT-LCD與半導體之製程極為類似，技術層次高而複雜，不論是上游設備、各種材料與零件的特性、面板製程與模組組裝技術，仍在持續開發改進中。且其每一畫素是單獨驅動，欲達到高良率是很大的挑戰，後進廠商要如何突破智慧財產權（IPR）的問題，是一個很大的考驗。

三、產品生命週期短及生產線技術更替速度快

TFT-LCD產業之另一個重要的特性即為產品生命週期短及生產線技術之更替速度快，且有加速之趨勢。廠商必須持續投入大量之研發費用，以研發出新技術，創造另一波競爭障礙，並因應市場趨勢，提昇產品之附加價值，加強服務品質，降低單位生產成本，提高產品品質，以維持競爭力。

另外根據李秀玉(1999)所歸納之TFT-LCD產業特性，有以下現象：

一、產業結構的完整性有利於競爭優勢之塑造

TFT-LCD之產業競爭力取決於上游材料及關鍵零組件之高自製率、中游面板之技術能力，及下游應用市場之多樣性三者之健全發展。其中材料成本占面板總成本的60%以上，因此，國內TFT-LCD產業欲達到成本與日商或韓商競爭之局面，必須提高國內面板製造廠之上游材料自製能力的提高，若材料完全依賴進口，成本即缺乏競爭力。

二、產品良率、品質與尺寸為競爭之關鍵因素

另外，良率、品質亦是影響成本之關鍵因素，成本計算時，良率會用來作為調整之用。此外，面板尺寸之選擇亦決定企業獲利能力之考量因素之一。一旦所有面板製造商皆將所有目標尺寸定為相同時，市場上同一尺寸之量多價跌的現象自然就會出現，因此廠商於選擇切割尺寸方面，不得不多蒐集競爭對手的資料，以做為決策之依據。

由以上的產業特性敘述中得知，TFT-LCD是一個非常重視技術層面的產業，雖然進入障礙大，但若是不注重自身產品品質的話，退出產業所要付出的成本是相當高的。在此情形之下，想要繼續生存就要能提供好的產品，因此提升良率就成為現在還在產業中的廠商最迫切需要解決的問題，

而國內相對於韓國廠商劣勢就是在於尚未達到整體產業垂直整合的地步，許多關鍵技術及零組件仍是掌握在別人的手中，因此，本研究希望提供一種參考手法，幫助國內 TFT-LCD 廠商能快速提升產品品質，也能提升整體競爭力。

2.1.3 TFT-LCD之相關研究

有關於TFT-LCD產業的研究大致可分為兩部分，其一為經營策略方面的探討，其方法從環境分析到廠商的策略選擇提供廠商在經營上的思維，可知TFT-LCD廠商若是想要在市場中占有一席之地，除了基本的經營手法之外，尚需其他多方面的配合，像是服務、配銷等，而本研究的主題為品質問題，在張嘉麟(2001)的研究中指出品質是TFT-LCD廠商所必備的核心能力之一，另外在朱正中(2005)的研究中指出垂直整合程度不同確實會對其營業利潤率的表現有不同的影響，表現在ROE上也會出現顯著的差異，台灣廠商目前最大的劣勢就是在於垂直整合程度不及日韓，若要提升競爭力必定要從品質問題方面著手;另一大項為製造程序方面的探究，與本研究所要探討的主題較為相似，在此項中，過去文獻大多著重於生產排程方面之探討，排程的工作涉及生產製程中有限資源的合理配置，以確保生產作業能儘量如期完成。是故排程計劃不僅需要有效率地解決各種資源分配所造成的衝突，同時也要解決各種生產限制條件的衝突，以使整個生產系統能夠平穩的運作，也是廠商欲改善品質不可或缺的改善工作之一(Garcia-Fornes et al.,1996)(Rodammer&White,1988)，以表2.1說明製造程序方面之文獻整理:

表2.1 TFT-LCD在製造程序方面文獻

作者(年代)	使用方法	內容及結果
李俊昇 (2003)	限制理論 (TOC)	發展一同步化集批法 (synchronized batching heuristic)，來解決Cell製程中批量生產派工方法，目的減少產能損失及縮短生產週期時間。
謝仲為 (2002)	先進規劃與排程系統 (APS)	解決TFT-LCD 產業在多廠 (FAB) 的各自不同限制下，同步協調各廠的排程計劃，以提昇各廠的產能利用率。
Jenog et.al (2001)	啟發方法 (heuristics)	解決Cell製程的排程規劃，目的求出最小的平均流程時間

作者(年代)	使用方法	內容及結果
Shin and Leon (2004)	MULTIFIT和禁忌搜尋 (tabu search)	發展出平行機台排程方法，解決TFT-LCD模組製程機台換線的問題
Toba (2004)	模擬系統	針對線上每一個批 (lot) 的在製品 (work in process, WIP) 進行排程控制，目的在不影響生產產能下，減少WIP的等待時間。

(資料來源:本研究整理)

2.2 中小尺寸TFT-LCD製造程序與品質問題

2.2.1 TFT-LCD 製造流程

TFT-LCD在中小尺寸與大尺寸之間生產製造技術別無差異，主要為結合半導體產業、化學材料產業及光電產業之製造技術，可分為：(1) Array製程，(2) Cell製程，(3) Module製程

1. TFT陣列基板段 (Array段)

陣列基板電路製程與半導體製程極為類似，首先須先清洗無鹼玻璃基板，再進行多層薄膜成膜工程(Thin Film Process)製作薄膜電晶體和透明電極層，而後再經光阻塗佈、曝光、顯像、蝕刻及光阻剝離等程序，反覆多次即可逐層堆疊出所需之線路與元件，形成薄膜電晶體TFT(Thin film Transistor)之陣列基板。目前由於台灣LCD產業之薄膜製程相關人才大都來自半導體產業，經驗豐富且技術成熟，故一般而言，製造良率皆可達95%以上。

2. 面板組裝段(Cell段)

TFT-LCD之面板組裝流程，首先將洗淨後的彩色濾光片基板與TFT陣列基板塗佈上配向膜塗液，並摩擦定向。然後在TFT陣列基板四周塗上框膠(Seal)，並散佈5~10 μ m大小之間隔物(Spacer)於其上作支撐點，再將陣列基板與彩色濾光片基板組合，以框膠封合形成空的Cell。接下之液晶注入工程有兩種製作方式，可先將此空的Cell之基板裁切斷裂片，取最終顯示器產品所需尺寸大小，經檢查工程後，再將液晶材料以真空方式注入並加以封合；或者先灌入液晶，進行裁切斷片工程後再封合。這兩種方式所需之製作時間不同，不但會影響總良率也會造成Throughput不同。最後於外層貼

上位相差板及偏光板後，即完成TFT-LCD面板之組裝。面板組裝段工程是目前影響TFT-LCD 面板良率最主要的工程段，除了不同材質之零組件組合技術較困難外，也需要豐富經驗之累積，目前此階段之良率尚達不到85%，未來仍有待提升。

3. 模組組裝(Module段)

最後將組裝完成的TFT-LCD面板與驅動IC(Drive IC)、印刷電路板(PCB)連接，並裝上背光源(Backlight Module)及固定框架(Frame)即完成液晶顯示器模組組裝。模組組裝(Module)段為目前TFT-LCD廠人力需求最多的部分，雖製程並不困難，但因無法自動化作業，必須用人工加工完成，所須人事成本較高，因此也有公司採外包作業，或轉至人力成本較低之大陸設廠，一般而言，總體良率已幾乎可達99%。

2.2.2 TFT-LCD品質檢測技術

本研究針對台灣地區某家中小尺寸TFT-LCD廠的生產製造流程進行分析，採用的是該公司的品質檢測資料。TFT-LCD檢測的方法主要分為光學檢測、電器檢測以及人工目視檢測。就目前TFT-LCD自動光學檢測的技術，已經順利的被導入在陣列製程及彩色濾光片製程上，並且應用在單層薄膜沉積後的瑕疵、玻璃基板微粒污染以及彩色濾光片上的異物、突起瑕疵、色彩不均、白色脫落等簡單的檢測，但是相較於大量使用人工檢測的模組檢測工程來說，如何將自動光學檢測的技術應用在模組點燈並且實際導入生產線上的最終畫質檢測上，是一個極待克服之問題。現今市場對於TFT-LCD的需求日益增大，且光電廠又得面對變化迅速之市場需求，故產品的品質檢測也逐漸被重視，但目前光電廠中最需解決的問題皆出現在Cell段及Module段製程品質，所以瑕疵檢測就格外的重要且為關鍵。(陳響亮&張頌榮,2005)

為因應未來檢測設備對於新製程使用與精密度要求的提升，自動光學檢測(Automated Optical Inspection; AOI)設備，已被廣泛用在目前TFT-LCD相關製程中，可用來取代傳統速率慢、人工成本高、檢測良率低、缺乏可重複性的人工視覺檢測設備(Manual Vision Inspection, MVI)。而電氣檢查以陣列測試器高速檢查TFT功能或周邊線路的電氣特性，檢出瑕疵的關連性比光學檢查更佳。目視檢查則用人眼來完成檢查的功能，如顯微鏡的表面檢查或面板/模組的最終畫質檢查，目前仍以受過訓練的人員進行檢查為

主；但為了判定的定量化與節省人力，自動面板檢查裝置的導入亦持續進行中。

根據製程可將TFT-LCD製程設備分為陣列（ARRAY）設備、面板組裝（CELL）設備、模組構裝（MODULE）設備、檢測修整設備與自動化設備等五大類。其中，檢測修整設備可依性質不同而分成四大類，如表2.2所示：

表2.2 TFT-LCD檢測設備分類

類別	裝置
檢查設備	<ul style="list-style-type: none"> ● 玻璃基板檢查設備 ● CF 檢查設備 ● TFT 圖案檢查設備 ● 陣列測試設備 ● 配向膜檢查設備 ● 巨觀檢查機 ● 點燈測試設備
量測設備	<ul style="list-style-type: none"> ● 顯微檢查機 ● 膜質/膜厚測定 ● 阻抗測定 ● 精密座標量床
修整設備	<ul style="list-style-type: none"> ● 雷射修整機
評價設備	<ul style="list-style-type: none"> ● 色彩、視角、應答速度 ● 環境、機械、外觀

（資料來源：金屬中心ITIS計畫整理（2006/03））

在TFT-LCD檢測工程中，若依目的則分成製程檢測與產品檢測兩種，如表2.3所示：

表2.3 檢測工程的概要

目的	檢測環境	檢測形式	檢測方法	製程項目
製程檢測	各製程中	透過製程的成果來檢出各製程的異常，藉由工程單的回報進行安定化和監控，透過監控模式切換到裝置內的參數管理，期望階段性縮減檢測工程。	光學檢測	素玻璃基板、Spacer、彩色濾光片、曝光/顯影
			人工目視檢測	配向膜整理、Spacer 封口、貼合組立、TAB 接線、PCB 接線
產品檢測	各製程結束時	全數檢查為原則，依據顧客要求的出貨檢查規格作為判定基準，大多是以人工目視檢測進行該製程最終的產品合格判定。	電器檢查	Array 最終檢測
			人工目視檢測	Cell、Module 最終檢測

(資料來源:本研究整理)

2.2.3 TFT-LCD 品質問題及相關研究

TFT-LCD是屬於高精密技術的產業，所以，掌控制程不良率的變動是相當重要的工作。影響製程不良率（或良率）的變數有很多，吳文智(2001)研究發現有下列幾項可能都會有影響：

1. 製程技術的不同：在玻璃基板尺寸逐漸增大的情況下，在前段和後段製程的處理難度上明顯提升不少，因為基板的面積大會產生曝光、顯影與蝕刻的均勻性問題，以及增加面板的切割困難度。所以，採用的製程技術不同，不良率就可能產生差異。
2. 產品尺寸的組合：當每批所生產的產品尺寸固定不變，且其他因素皆固

定不變時，其不良率可能不會有太大的改變，但是當產品尺寸需求改變時，或許會使原來的不良率發生變化。

3. 原物料的品質穩定度：生產的物料品質若能保持良好的水準，其影響產品製造良率的變動程度可能會比參差不齊的物料品質來的低。
4. 作業流程的改善：利用之前經驗，修改有助於降低不良率之作業流程，將可使整體良率提高。
5. 工具設計的改善：更正不良的工具設計或簡化其使用方式，可能會影響到不良率的發生機率，但其效應大小要視實際的生產情況而定。
6. 作業人員的訓練與流動率：現場作業人員若有經過工作訓練，對作業的熟練度會有所助益，且其生產良率的變動程度應會比未受過訓練來的小。若是作業人員的流動率高，不熟練工作操作或流程的新員工加入，可能會使原先已平穩之良率產生變化。
7. 生產設備的更動：汰換舊設備或添增新設備，須適應整體的生產情況，所以可能會改變原先的不良率發生機率。

在LCD面板組裝製程中，TFT陣列基板經過切割、研磨、配向膜塗佈、烘烤、配向處理、間隙物散佈等加工過程，其中配向膜塗佈厚度是否均勻、烘烤後厚度是否均勻、間隙物散佈密度是否均勻，在後序模組組裝與功能測試時，由於塗佈厚度不均勻或是間隙物密度不均勻，可能會造成模組無法組裝或是面板顯示時會有色差的情況，所以配向膜塗佈厚度是否均勻、烘烤後厚度是否均勻與間隙物散佈密度是否均勻是三個重要的品質特性(陳坤盛等,2004)。

雖然已經有學者提出以機器視覺的方式來判定品質，但目前TFT-LCD產業中的品質判定方法大部分還是採用專業訓練人員以裸眼的方式檢測，但是此方法往往在有瑕疵產生時需要從頭判定原因，耗時又耗力，因此本研究將以上述關鍵品質特性中面板組裝段(CELL)做為資料來源及分析的對象，找出能快速判斷造成品質不良問題的因素，以利檢測人員的判定速度及精準度。在TFT-LCD產業中，有關品質的相關研究如下表：

表2.4 TFT-LCD有關品質研究相關文獻

作者 (年代)	使用方法	內容及結果
呂俞德 (2005)	凝視追蹤系統與 卡片分類方法	訓練階段有助於改善搜尋形式、縮短搜尋時間與判定時間及增加判別面板缺陷之正確率，且專家大多照著規範行事，在一些缺陷檢驗中衍生出自己的搜尋策略。因此，有一些顯而易見的缺陷反而花費較多時間檢驗。
柯長泰 (2004)	Lingo、Opt Quest 等最佳化工具	求解Cell與Array段的投料時間、品項與數量並評估求解效果。且為節省求解時間，提出一簡易的FB演算法(Forward-Backward algorithm)
鄒民欽 (2002)	品管七大手法及 田口式品質工程	透過實際的參與，提出所觀察到的現象及整理出結論，對生產製造過程提出改善建議。
陳志忠 (2002)	CCD、影像擷取 卡、光學尺、線 性馬達移動平台	針對LCD面板的畫素瑕疵與亮度均一性進行判斷，其中包含量測定位、LCD面板取像、影像處理與自行發展的十字標記數學演算法等技術
曾彥馨 (2003)	傅立葉轉換	空間灰階影像週期性特徵有效地呈現於傅立葉頻譜影像中特性
錢志豪 (2003)	進階統計製程管 制方法、蟻行演 算法與倒傳遞類 神經網路(BPN)	經由研究分析與系統實作建立一套TFT-LCD色彩顯示之品質保證模式與自動化檢測程序

(資料來源:本研究整理)

由以上搜集的文獻可以發現，過去在研究TFT-LCD相關的品質問題時，利用了各種手法去改善製造時的流程和方法，或是建立一套可以依循的標準，可見TFT-LCD的製程改善是可以在不斷的實驗和研究中達成的，本研究試圖找尋判斷品質問題的法則，以利廠商在發現品質問題時能依據這些法則快速找出根源並改善之，提升製造良率。

2.3 資料採礦

2.3.1 資料採礦之定義及程序

拜科技進步及網路資料傳輸速度突飛猛進之賜，資料採礦(Data Mining)

近來已被大量運用為分析的技巧及工具，其定義有許多人提出，Fayyad(1990)認為資料採礦就是從大量資料中萃取出有效的、嶄新的，可能會被理解且最終有重要功效的過程;Han and Kamber (2004)提到資料採礦是從大量資料中挖掘或萃取出知識的一個程序，並利用這程序發掘出有用的資訊;而Richard J. and Michael W(2003)則認為資料採礦是一種從整個資料庫，利用一種或多種電腦技術去分析或擷取知識的過程。

簡單的來說，資料採礦就是從一大堆的資料庫中找出相關的模式，自動萃取出可預測的資訊，並利用這一些資訊輔助企業做出重要的決策。但這個概念並非單純只是統計學的迴歸分析以及資料庫管理系統，因為前者必須先有基本假設存在，且還需要處理大量資料的能力；而後者則是無法對資料作更進一步的分析。唯有資料採礦則是利用完善的統計和機器學習(Machine Learning)技術來建立自動預測及找尋規則的模型，同時還能結合商業資料倉儲(Commercial Data Warehouse)，發展出一套非常具有價值的商業用途模式，因此本研究決定採用資料採礦相關辦法對於TFT-LCD製造品質做有系統的分析，期能找出有用的資訊。

在資料採礦流程方面，Fayyad(1996)曾就資料挖掘流程的重點作深入的描述，他認為資料採礦程序包含八個步驟分別為：瞭解資料、獲取相關知識、資料整合與檢核、資料清理、發展模型、進行資料採礦、測試與檢核模型與結果解釋與應用。如圖2.4所示：

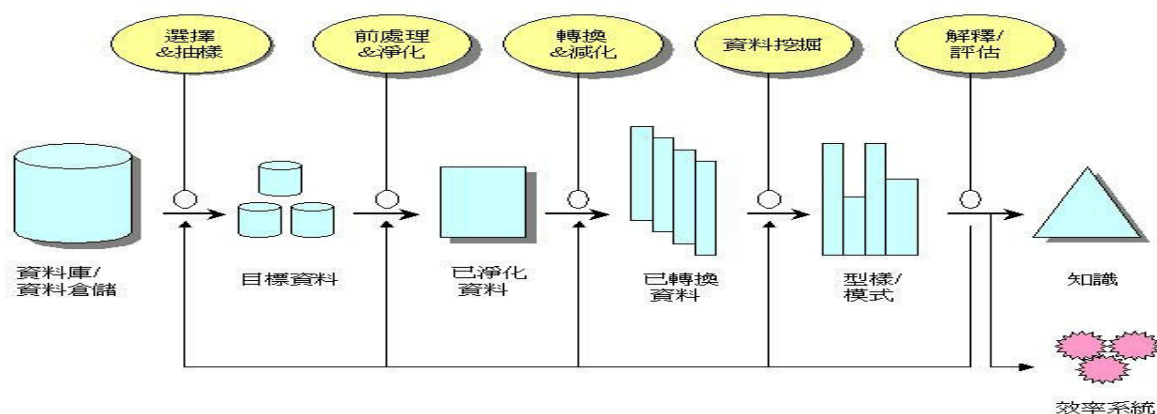


圖 2.4 資料採礦流程

Han and Kamber(2000)也曾提出類似的資料採礦流程，如圖2.5所示：

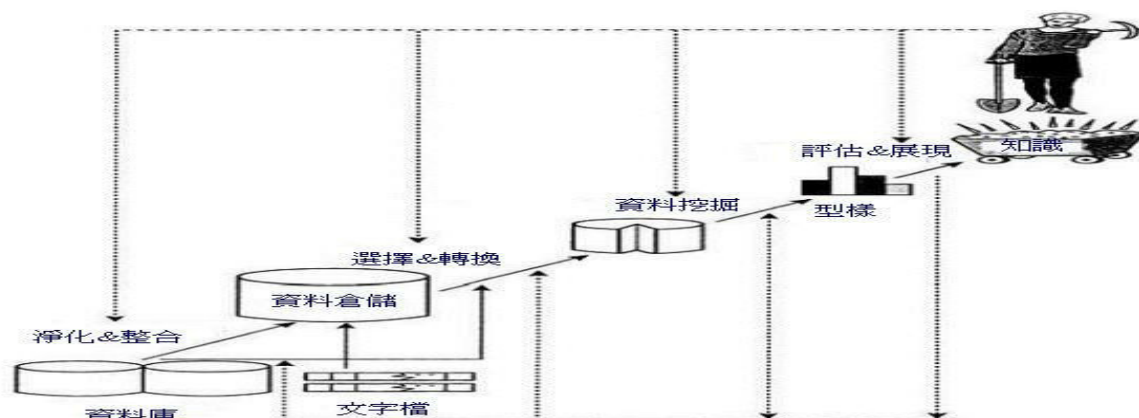


圖 2.5 資料採礦的重要步驟

另外其他學者也提出了一些關於資料採礦流程的見解：

1. Cabena (Cabena P. and Hadjinian P. et al 1998)：認為資料採礦程序包含下面五個階段：1、清楚明確的定義商業問題與挑戰2、資料準備包含：資料定義、資料擷取、資料品質檢驗、資料轉換3、選擇演算法建置模型4、解釋與評估模型5、以商業的眼光去洞悉資料採礦分析的結果並將其應用於組織的商業行為與資訊系統。
2. Kennedy (Kennedy R. and Lee Y. et al.,1998)：認為資料採礦程序為下列八個階段：1、定義問題 2、蒐集資料3、準備資料4、資料前置作業5、選擇演算法6、選擇演算法訓練參數7、訓練與測試模型8、評估模型。
3. Chapman (Chapman P. & Clinton J. et al.,1999)：提出CRISP-DM資料採礦程序模型，認為資料採礦程序包含六大階段分別為：1、產業認知2、資料透視3、資料準備4、模型建立5、成效評估6、規劃部署。
4. Kantardzic (Kantardzic ,2003)：認為程序由五個階段所組成包括1、詳述問題2、蒐集資料3、進行資料前置程序4、模型評估5、模型說明與結果描述。

2.3.2 資料採礦相關技術及功能

資料採礦的實際應用功能方面，目前較常被討論的有下列幾種(Elmasri and Navathe , 2000)：

1.分類模式(Classification)：

是指將資料依據事先定義的特徵屬性進行分類處理的工作。亦即根據

一些已知分類資料的變數屬性數值來進行運算，在得到各分類之相關規則後，再將未知分類的資料帶入規則以得到最後之分類結果。而這些使用者所運用來建立特徵的資料可能是來自於現有的歷史性資料，或是針對一個完整資料庫進行部份取樣(sampling)後的結果。

2. 分群模式(Clustering)：

所謂的分群模式是根據資料本身的特性而將資料進行分組的一個程序。通常在經過不斷的修正過程後，分群的特性會在群族本身的特徵上顯現出來；換言之，分群的目的是要將群組之間的差異辨識出來，並進一步指出個別組內的相似性。所以，事實上Clustering與Classification是不同的，其差異乃在於分析之前分析者並不知道會以何種方式或根據來進行分類。

3. 迴歸模式(Regression)：

乃是在瞭解及建立預測變數(X)與準則變數(Y)間的關係。其主要是針對現有資料進行運算，並利用預測變數的數值來預測準則變數的結果，進而建立預測模型。

4. 時間數列分析(Time-series Analysis)：

與迴歸模式的功能相似，兩者最大差異在於時間數列所分析的數值都與時間有關，譬如時間的週期性、季節性等，主要目的是預測未來趨勢的相關資訊。

5. 關聯式分析(Association)：

常被應用在超級市場的銷售物品上，其基本目的是希望在多屬性(Attributes)或多品項組合(Item Sets)的關聯探討中，尋找各種銷售品項(items)彼此之間在銷售情形的關聯性。主要著重點則在於關聯性可信度(Support-Confidence)的架構。

6. 順序分析(Sequential Pattern Analysis)：

此模式是針對資料的順序關係進行採礦，例如在時間上的順序、行為上的順序等，而主要是希望能找出趨勢或順序上的分析資訊。

本研究主要運用的手法為分類與分群相關方法，以及改良技術的決策樹方法，以表 2.5 詳加說明之：

表2.5 分類與分群之說明與技術

Classification	<ul style="list-style-type: none"> ● 根據一些變數的數值做計算，再依結果作分類。 ● 用一些根據歷史經驗已經分類好的資料來研究它們的特徵，然後再根據這些特徵對其他未經分類或是新的資料做預測。 	
Clustering	<ul style="list-style-type: none"> ● 將資料分群，其目的在於將群間的差異找出來，同時也將群內成員的相似性找出來。 ● 與Classification不同的是，分析前並不知道會以何種方式或根據來分類，所以必須要配合專業領域知識來解讀這些分群的意義。 	
理論技術	傳統統計技術	<ol style="list-style-type: none"> 1. 因素分析 (Factor Analysis) —精簡變數 2. 判別分析 (Discriminant Analysis) —分類 3. 分群分析 (Cluster Analysis) —區隔群體
	改良技術	<p>決策樹 (Decision Tree) —用樹枝狀展現資料受各變數的影響情形之預測模型，根據對目標變數產生之效應不同而建構分類的規則，多運用在對顧客資料的區隔分析上，本研究將利用此方法試圖找出隱藏於不良資料中的可能肇因。</p>

(資料來源: 謝邦昌(2003), 本研究整理)

資料採礦目前已被應用在許多領域，例如行銷、銀行、通訊產業，製造等方面，其中在製造業方面已廣泛的被使用在控制及排程技術生產程序，例如LTV Steel Corp.為全美第三大的鋼鐵公司，該公司使用資料採礦手法來偵測製造過程中潛在的品質問題，使得他們的不良品減少了99%(洪煌龍,2004)，由此可見使用資料採礦的方法對於產品的品質問題做相關探討可以大幅減少不良品發生的機率，對於管理方面的效率相對也提高許多，Andrew(2000)在半導體的製程中，以資料探勘中的決策樹方法探討晶圓磨平、圓(roll-off)的問題上，在他的研究中指出「在晶圓的製程過程中，如果品質不良的話，則必須多透過一個磨平晶圓的過程，而此過程將會造成過多的生產浪費，所以要如何在常造成品質不良之問題製程中找出成因」，他的研究應用C4.5決策樹演算方法找出影響此問題最嚴重的成因規則，進而幫助工程師配合其專業知識針對此成因進行改善工作。而Chi(2001)等學者亦運用資料探勘中的C4.5決策樹方法解決相關問題，他的研究是以可攜式電子產子之抗摔程度做規則化之成因做因子的擷取及相關探討，並且將決策樹所擷取出的參數規則提供專家、工程師做改善決策的諮詢及參考。

國內學者林鼎浩(2000)同樣是針對半導體不良率成因問題做出討論，他利用ART類神經網路手法將晶圓的受損特徵加以分類，再輔以k-W統計檢定決策樹分析方法找出影響良率最嚴重的參數規則，並進而將所得的資訊提供給工程師做為解決問題、提升良率的參考依據。李培瑞(2001)表示在傳統統計、品管均無法掌握的大量生產製程的多維度變數資料，他應用C4.5決策樹演算法，結合了SOM類神經網路方法，找出影響良率最大之製程參數規則，其作法是將數百個維度的製程資料先濃縮成重要的18個主要影響參數做為輸入值，套用SOM類神經網路找出良率較低的分群，並且運用C4.5決策樹方法於以低良率群體為分群，並求出問題成因之規則化語意，而工程師則可以透過此規則化語意，將參數之間的相關敏感度有更清楚的了解，以提供現場工程師做為進而提升晶圓良率的決策參考。

另外Andrew(2001)應用資料探勘的手法於PCB製程問題上，他以PCB組裝過程中的焊球(solder balls)的產生為研究範疇，亦利用決策樹方法來取代傳統統計方法中無法有效以實驗設計方法求得的問題產生因子，將相關之製程參數做為輸入項，並求得三項參數規則，而經由工程師以其專業知識做為比對後，他的研究中發現其中的二項規則符合工程師猜測的可能「知識」，而另一個參數規則的發現經由現場工程師確認後，成為他的研究中最大的貢獻所在。Dan(2002)則將資料探勘的手法用來尋求改善半導體製造業中的微粒(particles)掃除製程，他也利用C4.5決策樹方法將製程中的製程參數(包括雷射密度等等)規則化，最後在將參數規則化結果提供工程師做為改善製程的知識。

2.3.3 資料採礦應用於TFT-LCD之相關研究

由於TFT-LCD產業於世界上的主要生產加工組裝乃至新進製程研發地點的主力都在亞洲地區，因此相關研究不論是產業策略或是製程改善的研究，大多是由亞洲地區的國家學者進行研究，相對的歐美地區學者對於此類的相關研究可說是少之又少，因此舉出國內多個以TFT-LCD為背景產業，並利用各種資料採礦方法做為分析工具的研究，如表2.6。

從表2.6可以看出，過去已有些對於TFT-LCD製程品質的探討，並且也運用了許多不同的資料採礦方法來作分析，得到了一些用以改進製造過程，提升產品良率的手法。本研究擬進一步針對TFT-LCD製造過程中的關

鍵製程，也是目前良率最低的面板組裝(CELL)製程的不良資料作成因的決策樹分析，希望透過這樣的研究能讓廠商能更直接的判斷是哪些原因造成最後的不良現象，進而縮短整體檢測時間，也能提升生產效率。但在過去研究指出，若是先對資料做分群再進行決策樹分析，會比直接使用決策樹分析有更高的正確率(簡禎富等,2003)(鄭惠珊&鄭景俗,2007)，因此本研究使用的資料採礦方法為先對目標資料做分群(cluster)，再依據分群資料建立造成不良品質原因的決策樹分析(decision tree)，對於方法的架構與流程將於第三章中說明。

表2.6 資料採礦應用於TFT-LCD之相關研究

作者(年代)	使用方法	內容及結果
洪瑜君 (2006)	決策樹、叢聚分析法	找出最有可能的產生缺陷的製程，針對特定製程所產出之不良原因資料，再將不良缺陷資料分群，最後以視覺化技術，分析是否具有缺陷群聚現象產生
王人星 (2004)	自組織映射網路	模擬真實的TFT-LCD廠模型為基礎，於實驗TFT-LCD廠中的製造批量大小，來觀察其變化對於績效指標的影響，運用資料挖礦的方法來找出在製程中的關鍵機台，並且整理有效的資訊蒐集有效的模式
郭信宏 (2007)	羅吉斯迴歸、類神經網路、C5.0決策樹，交叉分析	得出不良率檢出準確度最高、耗費人力最少之分類法則，成功達到人力成本減少及品質提升之目的，發揮資料探勘幫助企業追求利潤、提高顧客滿意度之目標。
周雅君& 鄭春生 (2006)	CHAID、Exhaustive CHAID、CART、 QUEST	供應商品質管理 (supplier quality management, SQM) 可以藉由此套品質異常診斷系統快速分類，進行機器設備與參數設定最佳化與特殊檢驗手法的搭配，並水平展開到所有供應商管理系統，降低異常重複發生機率
黃瓊芬 (2003)	Apriori algorithm	找出存在於TFT-LCD產業上下游間股價之特殊現象，並改進Apriori algorithm使其適用於跨交易日離散型非連續性之資料
莊智偉 (2007)	無母數分析K-W檢定、Box-Plot、決策樹	針對適用於液晶面板製造工廠之產品品質變異成因分析模式，建構產品品質異常事故診斷系統來協助推論異常事件根本原因

作者(年代)	使用方法	內容及結果
Yu Zhang& Jian Zhang (2005)	模糊類神經(fuzzy neural network)	研究發現模糊類神經網路對於解決顯示器亮度不均勻造成各種痕跡的現象(MURA)的問題是非常有用的工具
Tseng& Tsai (2006)	獨立成分分析 (Independent Component Analysis,ICA)	針對TFT面板之製程參數變異偵測，偵測對象為總斜度變異量(Total Pitch variance, TP)，透過總斜度變異量的監控可以有效回饋生產線及時且快速之製程變異資訊，避免大量不良品的產生。

(資料來源:本研究整理)

2.4 本章小結

由本章的探討可以發現，TFT-LCD的品質問題是影響其在市場中競爭優勢的最大因素，因此本研究選擇最關鍵的製程且為目前良率最低的面板組裝(CELL)製程來做為分析對象，而由前述文獻中可以得知，目前此製程的檢測方式是大多還是使用人工目視的檢測方法，可能會有檢查速度過慢或是對於問題根源判斷錯誤的情形產生，因此本研究希望找出一個可以快速檢驗並找出真正問題根源的方式，以減少不良品的產出。

而從資料採礦理論集資料採礦類型的探討中，可以知道資料採礦主要是用來幫助決策人員或使用人員快速找到問題的來源，從過去的歷史資料中，經過資料採礦方法的演算歸納，推導出強而有力的法則。本研究即是利用此一方法中的決策樹歸納學習法，期望能夠建立明確且易於了解的決策模型，將複雜的決策判斷程序簡化，提供較易於解釋的答案(Safavian and Landgrebe, 1991)，以利檢測人員可以更快速且正確的對問題根源做判斷，而在進行此方法之前，也先對資料進行分群以求得更正確的結果。

第三章 研究方法

在本章中，首先針對資料採礦中的類神經網路方法做介紹，包括運作方式及不同學習方式之探討，接著對分群方法中的自我組織映射圖網路及二階段分群法進行運作方式說明，最後再針對本研究用來找出法則的決策樹方法加以說明，本研究希望能透過這些方法得到能迅速找出品質問題根源的法則。

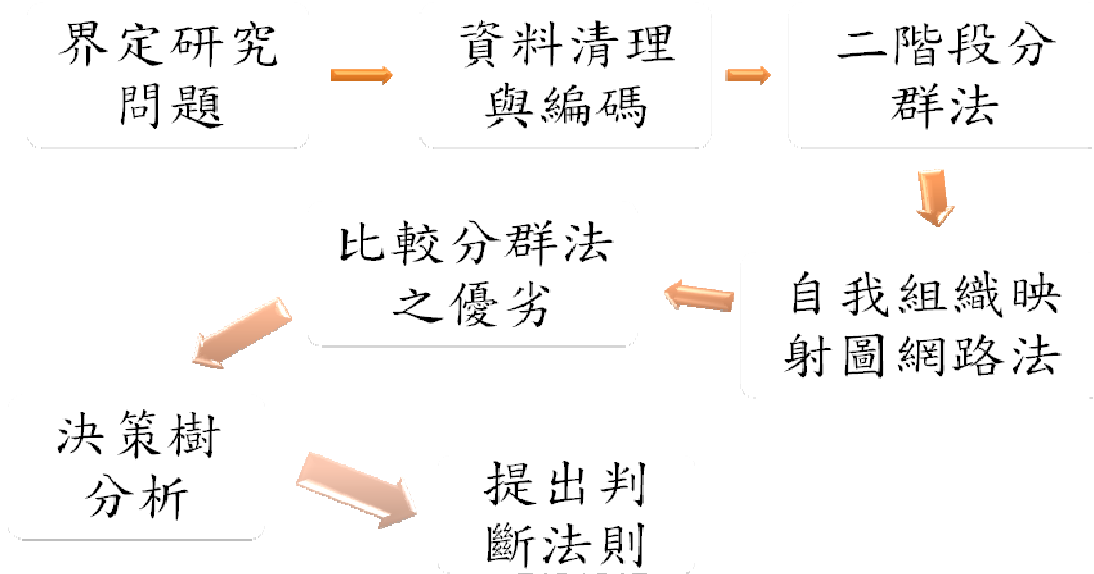


圖 3.1 研究方法架構圖

(資料來源:本研究整理)

3.1 類神經網路

3.1.1 類神經網路之定義及原理

類神經網路是人類為了模仿生物神經網路的資訊處理系統所設計出來的，其較精確的定義為：『類神經網路是一種計算系統，包括軟體與硬體，它使用大量簡單的相連人工神經元來模仿生物神經網路的能力。人工神經元是生物神經元的簡單模擬，它從外界環境或者其它人工神經元取得資訊，並加以非常簡單的運算，在輸出結果到外界環境或是其它人工神經元。』(Robert Hecht-Nielson ,1990)

當外界各種不同形式的資訊或能量訊號(例如：聲、光、電、熱等刺激)，透過感知器官轉換成電訊號之後，便會進入神經細胞中，而在神經元內此輸入的訊號是經由突觸(Synapse)間內部電位轉換後，透過樹突(Dendrites)傳

送至細胞本體(Soma)，再由軸突(Axon)傳送到樹突，成為下一個神經元的輸入訊號(圖3.1)。其中，突觸內的影响強度就是生物神經網路儲存資訊的地方，生物神經網路的學習就是在調節突觸的強度。由此可知，樹突是細胞的輸入路徑，透過細胞本體間許許多多聯絡的突觸來接受周圍細胞本體所傳出的訊號，那麼軸突就相當於細胞本體的輸出路徑。類神經網路是由許多人工神經元所組合而成的，其主要是要對生物神經網路作最簡單的模擬，因此人工神經元是最基本的類神經網路單元，稱為處理單元(Processing Element)或節點(Node)。

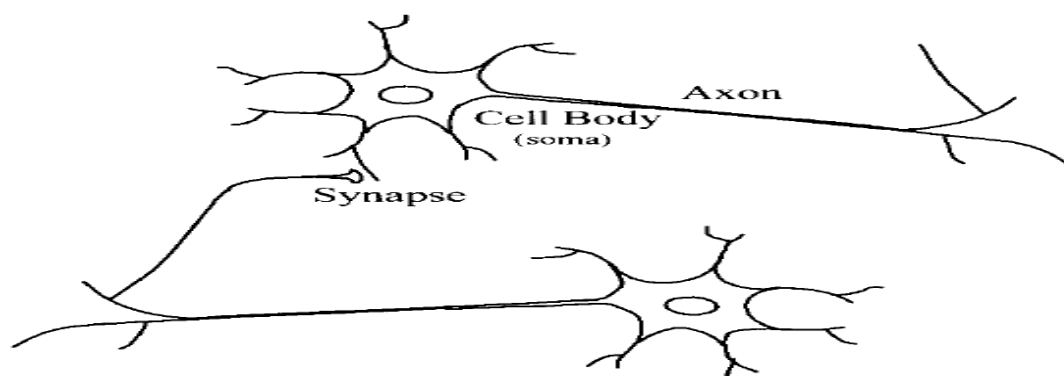


圖3.2生物神經元(資料來源:羅華強,2001)

3.1.2數學式

一般可用輸入值與其加權值乘積之和函數來表示：

$$Y_j = f(\sum_i w_{ji} x_i - b_j) \quad (3.2.1)$$

圖3.2為相關變數之功能，介於處理單元間的訊號傳遞路徑稱為連結(connection)，每個連結上均有一加權數值 w_{ji} ，用來表示第i個處理單元對第j個處理單元之影響強度。

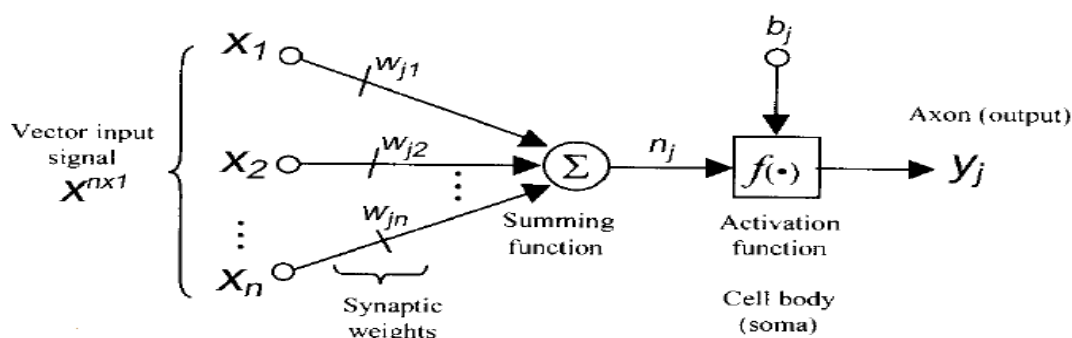


圖3.3人工神經元(資料來源:羅華強,2001)

3.1.3 網路的運作方式

基本上，類神經網路內部之運作分成二個過程：

1. 訓練過程—決定權重值(Weight)和偏權值(Bias)。
2. 模擬過程—決定預測輸出值或驗證網路準確度。

在訓練過程中，網路遵循所謂的學習法則(Learning rule)，各種類型的網路學習法則各異，而其目的即在修改網路的權重值和偏權值。一般採用訓練演算法(Training algorithm)，各種類型的網路訓練演算法也互異。

訓練過程將所需訓練的資料(Training data set)分成二類：

1. 訓練數據集—目的是修改權重值和偏權值。
2. 測試數據集—目的是驗證網路準確度。

所以，類神經網路的運作就是透過訓練，不斷的調整節點之間的權重值和偏權值，使網路計算出的輸出值趨近於目標輸出，讓網路能映射出一組正確的輸入—輸出關係模式。

3.1.4 學習法則—監督式學習與無監督式學習

學習法則的定義「用來修改網路權重值和偏權值的步驟，即是一種訓練演算法」，其可分成二大類：

1. 監督式學習—提供適當的網路行為訓練範例給學習法則，如 $\{(p_1, t_1), (p_2, t_2), (p_3, t_3), (p_4, t_4), \dots, (p_n, t_n)\}$ ，其中 p_n 是網路的輸入， t_n 是相對應正確的輸出(目標)。當輸入給網路時，網路的輸出就會與目標作比較，緊接著，學習法則被用來調整網路權重值和偏權值，目的是為了讓網路的輸出更接近於目標。
2. 無監督式學習—權重值和偏權值的修改只回應出網路的輸入，目標輸出值無從獲得。此種演算法大部分會執行叢集運作，其可以將輸入圖樣分成有限個數的種類，在向量量化中此種學習方式非常有用。

3.2 集群分析

集群技術(Clustering Technology)是一種非監督式的資料探勘方法，可以從龐大的資料中，找出資料的分布狀況並找出隱藏在資料內部特徵的意義，先將大量且雜亂的資料分成若干個集群，再針對不同的集群加以分析，如此便可簡化使用者在分析資料時的複雜度。而集群技術的主要目的在分析資料的內容，將性質相似的資料群聚在一起，而讓集群與集群間的資料相異性大。集群技術與傳統分類(Classification)最大的不同是，集群技術不

預先設定分類所代表的意義而把資料先以集群技術將性質相近的資料聚集成一群群的集群後，再分析定義各集群的意義。

在諸多集群方法中，目前最常被應用在不同研究領域的集群技術，包括模糊分群法、類神經網路分群法(SOM)和切割式分群法(K-means)，三者各有其適用的範圍(Vetter,1993)，但模糊分群法主要針對不明確屬性的資料進行分群，因此不在本研究討論範圍內。而SOM以網路拓撲中的空間概念進行鄰域的聚類移動，適用於本研究資料之分群；而K-means求解速度快並可將所有資料完全分割至集群，符合求解問題中必須完全配送所有資料的目標，因此將於下列詳細介紹SOM與K-means有關之二階段分群法之概念與相關應用。

在文獻中，曾經利用蒙地卡羅演算法，來比較類神經網路和非階層式的K-Means集群法做一個比較，結果是K-means方法有較高的分群正確率(Balakrishnan, 1996)。但是資料的錯誤率會影響K-means執行的效果，在資料無錯誤率的狀況下，所得的結果品質是最佳的，然而只要將錯誤資訊加入到資料中，就會降低結果的品質(Balakrishnan et al.,1994)。而在本研究的個案廠商中，TFT-LCD的品質檢測過程大多是以人工目視的方法來檢測，其資料正確率相對於用光學檢測低很多，因此必須要有兩種分群方法互為佐證。本研究將先以二階段分群法中的華德法決定分群數，再以K-means進行分群正確性的驗證，最後以K-means的結果與SOM做分群正確性的比較，並以較正確的資料來做問題的分析，而這部分將以決策樹方法進行，本節將說明本研究使用的三種方法及意義。

3.2.1 自組織映射網路

1.簡介

自組織映射圖(Self-Organizing Maps，以下簡稱SOM)是由T.Kohonen於1989年所提出之神經網路模式，其乃基於「競爭式學習」(Competitive Learning Methods)的一種神經網路，也是一種非監督式學習(Unsupervised Learning)的神經網路，在自組織特徵映射圖網路中，網路輸出層的神經元是以矩陣方式排列(Ulthsh&Siemon,1990)，相鄰之間均有連接，根據輸入向量(Input Vector)，互相以競爭方式取得到調整網路加權值機會，而最後之輸出層(Self-organizing Output Layer)的神經元會根據輸入向量之「特徵」(Feature)以有意義的「輸出結構」(Topological Structure)展現在輸出空間中，

由於所產生的輸出結構圖可藉由U-Matrix(Iivarinen,1994; Kraaijveld,1995)等技術映射到二維空間中呈現出來，以反應輸入向量本身的特徵，因此將此網路稱作為自組織特徵映射圖網路。

自組織映射網路其輸出處理單元會相互影響，當網路學習完畢後，其輸出處理單元相鄰近者會具有相似的功能，也就是具有相似的連結權數。SOM可將輸入值映射至輸出平面上的一個點，且相鄰近的點具有相似的功能，如圖3.4

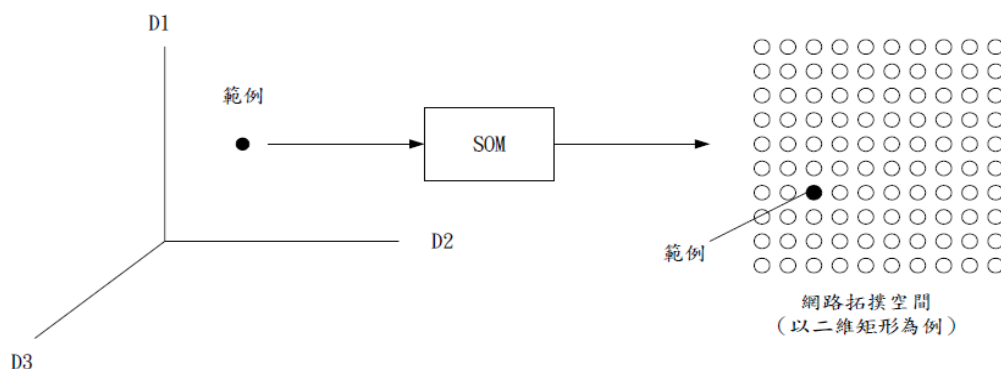


圖 3.4 SOM 的空間映射概念 (以三維樣本空間及二維輸出為例)

(資料來源：張華城(2000))

2.網路架構

無監督式學習的應用是從問題領域中取得訓練範例 (只有輸入變數值)，此研究將品質問題的基本資料作為輸入變數，透過訓練品質基本資料分類因子的內在分群規則，再應用於新的品質問題資料案例 (有輸入變數值，而需推論它與那些訓練資料範例屬同一聚類的應用)，如圖 3.5 示：

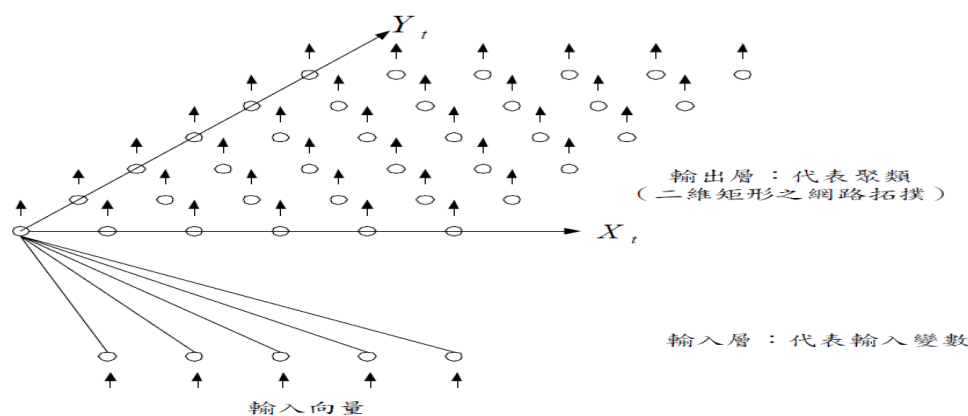


圖 3.5 SOM 架構 (只繪出一個輸出單元的連結)

(資料來源：葉怡成(2003))

自組織映射圖網路的重要概念，包括輸入層、輸出層、網路連結、網路拓撲、拓撲座標、鄰近區域、鄰近中心、鄰近半徑等。分別敘述如下(葉怡成, 2003)：

- (1)輸入層：用以表現網路的輸入變數，即訓練範例輸入向量，或稱特徵向量，其處理單元數依問題而定。使用線性轉換函數，即 $f(X)=x$ 。
- (2)輸出層：用以表現網路的輸出變數，即訓練範例的聚類，其處理單元數目依問題而定。和反傳遞神經網路最大差異在於自我組織神經網路有「網路拓撲」、「鄰近區域」的觀念。
- (3)網路連結：每個輸出層單元與輸入層處理相連連接的加權值所構成的向量，表示一個訓練範例對應樣本點聚類的形心座標。當網路學習完畢後，其處理單元相鄰近者會具有相似的連結加權值。
- (4)網路拓撲：自組織映射圖網路和一般類神經網路有一點重要差異，它的輸出層處理單元的相對位置具有重大的意義，而一般的網路架構模式則否。通常它的輸出層處理單元以二維的型態排列居多，形狀以矩形居多存在。
- (5)拓撲座標：意指標定一輸出層處理單元在網路拓撲中的位置座標，對於一個二維型態排列的網路拓撲，每一個輸出層處理單元具有一個二維拓撲座標。
- (6)鄰近區域：意指網路拓撲中，以某一輸出單元為中心的區域，稱此為一單元之鄰近區域，鄰近區域內的輸出處理單元會互相影響，且鄰近區域會在網路學習過程逐漸縮小。
- (7)鄰近中心：控制鄰近區域的中心位置之參數，意即一輸出層競爭優勝單元 (winner) 在網路拓撲座標。
- (8)鄰近半徑：控制鄰近區域大小的參數，一般採取拓撲座標橫軸單元數目相同，以 R 表示。

3.演算步驟

(1)步驟一:初始化

在進行SOM網路訓練前必須針對網路結構與權重向量進行初始化。在網路架構定義方面，基本上SOM的網路拓撲是由 i 個單元的集合所構成，並根據事先定義的固定拓撲型態來排列，最常使用的就是 $N * M$ 的二維網格。其次，每一個單元 i 都會被賦予與輸入資料相同

維度的權重向量 m_i ， $m_i \in R^n$ ，而權重向量(鏈結強度)可能有兩種方式決定，一是以隨機亂數指定，或以隨機策略如主成份分析(principle component analysis, PCA)來設定。值得注意的是，所有 i 個權重向量的初始值均應相異，且通常會加以正規化(normalize)成為長度為1的單位向量。

(2)步驟二:輸入範例特徵向量

針對時間/訓練循環，輸入向量 $x = (x_1, x_2, \dots, x_i)^T$ ，在此使用了一離散時間標記 t ，代表目前訓練重複次數。在每次訓練循環， $x(t)$ 均由輸入向量集合 R^n 中隨機抽出。

(3)步驟三:尋找優勝神經元

以最小歐基里德距離的方式尋找時間/訓練循環 t 之優勝單元 c (winner)，顯示其含有最高活動力。而對於隨機選擇的輸入向量 $x(t)$ 而言， $c(t)$ 則將顯出更高的活動力。換句話說， $c(t)$ 在特定輸入向量的未來表現皆會因為呈現較高的活動力而較為合適成為優勝單元。

一般而言，一個單元的活動力(作用值)是根據輸入向量(input pattern)與該單元的權重向量(鏈結強度)之間的歐幾里得距離(Euclidian distance)來衡量。換句話說，若有一單元其權重向量 m_i 與目前輸入向量 x 之間的歐幾里德距離為最小時，該單元便稱為優勝單元。因此，優勝單元 c 的選擇方式可用下列算式加以表示：

$$c(t) = \arg \min_i \|x(t) - m_i(t)\| \quad (3.2.2)$$

$c(t)$ 時間點 t 的優勝單元

$x(t)$ 時間點 t 隨機抽取的輸入向量

$m_i(t)$ 時間點 t , i 單元的權重向量

$m_c(t)$ 時間點 t , 優勝單元 c 的權重向量

$\|x(t) - m_c(t)\|$ 在時間點 t , 優勝單元 c 的活性(activity), 用來計算 Δm_i (鏈結強度修正量)

(4)步驟四:調整權重向量

適應行為(adaptation)發生在每次學習重複過程中，其執行方式是根據各別輸入向量與權重向量之差異量，朝向坡降 (gradual reduction) 方式進行學習。而至於適應調整的總次數是由學習速率 α 主導，其

亦會隨著時間過程而逐漸遞減。這種適應(調整)強度的漸減特性使得在學習初期有大量的適應步驟，並且權重向量亦需從隨機初始值逐漸調整轉向至輸入向量的實際需求(群心)。而在學習末期，微幅的適應讓權重向量在輸入空間宛如進行微調動作(如圖3.6所示)

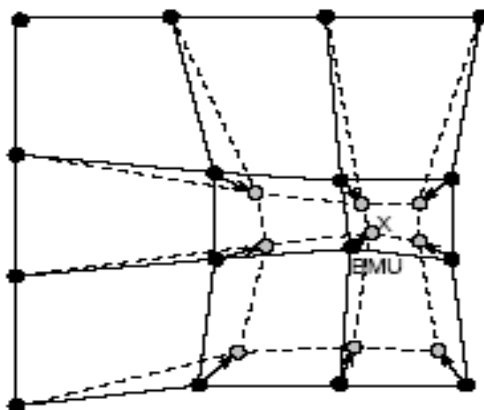


圖 3.6優勝單元與鄰近區域內權重修正圖(Vesanto,2000)

結合上述 SOM 的原理原則，我們可以寫出一如下列公式的學習法則來表示鄰近區域內單元之權重修正。在此使用離散時間標示 t 以代表目前學習循環次數。

$$m_i(t+1) = m_i(t) + \alpha(t) \cdot h_{ci}(t) [x(t) - m_i(t)] \quad (3.2.3)$$

t 表示目前學習循環次數

α 則代表隨時間變化的學習速率

h_{ci} 則表示隨時間變化的鄰近函數值

x 表示目前的輸入樣式(向量)

m_i 則代表分配給單元 i 的權重向量

(5)步驟五:返回步驟二，直到特徵映射圖形成後終止。

3.2.2 二階段分群法

在現今的統計學之中，要將樣本給劃分為幾個集群時，最常使用的一個程序就是集群分析方法(Cluster Analysis)，其中可分為階層式集群分析(Hierarchical Cluster Analysis)和非階層式集群分析(Nonhierarchical Cluster Analysis)，而兩階段方法就是結合了階層式與非階層式的優點，以求得最佳群數。分別說明如下：

- 1.階層式集群分析：主要是利用凝聚(agglomerative)與分離(divisive)的方式，將樣本的個體視為一群，然後將相近的個體合成同一群，依次的結合使得群組愈來愈少，最後所有個體結成一群，或是經由相反的程序，

將所有個體從同一群，一一分離出為個別的群數。此種方法最大的好處可以使決策者依據實際需要，選擇最適當的群數，作為區隔資料之依據。而群集間距離的判定有下列方法：

(1)最短距離法(nearest – neighbor method)

又稱為單一聯結法(single linkage method)，其採用群集間之最近距離，作為群集間之代表性距離。

$$d_{\min}(G_i, G_j) = \min_{x_i \in G_i, x_j \in G_j} d_{ij} \quad (3.2.4)$$

(2)最長距離法(furthest – neighbor method)

又稱完全聯結法(complete linkage method)，其與最短距離之聚類方式相反，其採用兩群集間最遠點之距離作為兩群集間之代表性距離。

$$d_{\max}(G_i, G_j) = \max_{x_i \in G_i, x_j \in G_j} d_{ij} \quad (3.2.5)$$

(3)群平均法(group – average method)

最短距離法及最長距離法各有其優缺點，觀念上，前者為空間之收縮，後者為空間之擴散。而群平均法又可稱為平均聯結法(average linkage method)，表示兩群間所有元素間距離之平均值。

$$d_{\text{ave}}(G_i, G_j) = \frac{1}{n_i n_j} \sum_{x_i \in G_i} \sum_{x_j \in G_j} d_{ij} \quad (3.2.6)$$

(4)華德法(Ward method)

華德法又稱離差平方和法(squares sum of dispersion method)和最小變異數法 (minimun variance method)，是由Ward (1963) 提出。此法先將每一樣本資料視為一群集，後再依序將群集合併，合併順序的安排是基於方差分析概念，以離差平方和為相似性測度。同群集樣本之間離差平方和最小，群集與群集間的離差平方和最大。假定已將 m 個變量的 n 個樣本分成 T 類， X_{ikt} 表第 t 類第 i 個樣本之第 k 個變量， n_t 表示第 t 類樣本之個數， \bar{X}_{kt} 表示第 t 類第 k 個變量的重心，則第 t 類中樣本之離差平方和為：

$$S_t = \sum_{i=1}^{n_t} \sum_{k=1}^m (X_{ikt} - \bar{X}_{kt})^2 \quad (3.2.7)$$

其中

$$\bar{X}_{kt} = \frac{1}{n_t} \sum_{i=1}^{n_t} (X_{ikt}) \quad (3.2.8)$$

T 個種類之離差平方和為：

$$S = \sum_{t=1}^T S_t = \sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{k=1}^m (X_{ikt} - \bar{X}_{kt})^2 \quad (3.2.9)$$

華德法在類和類合併時，是針對某兩類合併後，其 S 為其中增加最小者，予以合併之（高士傑,2001；黃俊英,2003；易任和葉惠中,1990）。本研究使用此演算法代表階層分群方法。

2.非階層式集群分析：非階層式集群分析中最常使用的方法為K-Means Cluster分群法，與其他的分群法最大的差異在於：若樣本的分群數為已知的情形之下，則K-Means Cluster分群法可以將所有的樣本分為指定的群數，而且K-Means Cluster分群法比較不容易受到不適當之屬性（構面）、異常點與不同的相似度（距離）的影響，因此本研究採用K-Means Cluster分群法與SOM的分群法做比較。

K-means又稱為Forgy's algorithm(Jain,2001)，其主要概念是以反覆疊代的動作，逐次降低目標函數值，直到目標函數不再變化，即代表分群結束。簡單來說，K-means演算法是一種將 n 筆輸入資料依照其資料特性分群為 K 類的分群演算法，其中 K 為一正整數，其演算法的目標在於求取各個輸入資料與其相對應分群中心點距離平方和之最小值。

主要的演算步驟如下：

步驟1：任意由 n 個資料物件中，隨機選取 k 個物件視為 k 個群聚的中心。

步驟2：對於所有 n 個物件，找其最近似的群心(以距離近者相似度較高)，然後將該物件歸到最近似的集群。

步驟3：計算目標函數(式3.2.10)，若未改變，表示沒有任何物件變換所屬群聚，可以結束分群，若有改變，則進行步驟4。

步驟4：重新計算新的集群中心點(群聚內各物件的平均值)，回到步驟

2。平方誤差標準(Square-Error Criterion)之計算如下：

$$E = \sum_{i=1}^k \sum_{p \in c_i} |p - m_i|^2 = 0 \quad (3.2.10)$$

(m_i 代表群聚 i 的中心， p 是群聚 i 內的物件， c_i 則代表群聚 i)

K-means群聚法簡單且有效，其分群結果佳、誤差率低，在屬性增加

時仍保持較低的錯誤率(Balakrishnam,1994)，但K-means在使用上的困擾為必須事先設定分群數目K，但一般較難從原始資料中判斷適當的群數，且在使用者主觀的認定並不斷測試下，容易造成客觀度不足的問題。因此發展出選取K-means初始群數的方式，分為四種(Pena et.al,1999):

(1)Anderberg隨機法：將資料透過隨機方式，分割成K個群組。

(2)Forgy方法：參考Anderberg的方法，由資料庫中隨機選取K個種子，藉著最近的種子分配到所代表的集群。

(3)Macqueen方法：由資料庫中隨機選取K個種子，依據其順序指派剩餘的種子到最接近群心的集群，經過每一個指派，所有群心重新計算完成。

(4)Kaufman方法：最初群數的決定來自連續代表種子的選取，直到K個集群都被找到。

3.兩階段方法：Sgarma(1996)建議如果要求的更精確的分類結果，可以同時使用分層法與非分層法。第一階段以華德法做分群，以決定群組個數，第二階段再以K-Means進行群集，以二階段方法目的是由於第一階段華德法是屬於分層式群集分析，當二個個體一旦被分在一群，則其後永遠就在同一群，而此時在第二階段所使用K-Means就是彌補此種缺點，以達到最佳組內同質、組間異質的群數。

此種分析的手法就是透過兩階段分群方法，修正了分層式集群法中對於集群分析一旦形成，就無法對併入不適當集群觀察值進行重新分群的缺點。同時也克服了非層級式集群法需事先決定集群數目與集群中心點的問題。

本研究除了採用前述的自我特徵映射圖網路(SOM)來做分群的動作之外，也因為若當SOM所產出的分群結果不是分界良好的集群時，會造成某些模糊地帶中的資料點無法明確歸屬於集群中，藉由K-means先行群聚可有效將所有資料點明確分割(Abidi&Ong,2000)，但也因需要事先決定集群總數和起始向量，就分群的實用性而言，似乎有所不足，因此將兩種方法對同一資料做一相同探究，可更真實並準確的呈現資料的原貌與邏輯，對於資料的解讀也更有幫助。

3.3 決策樹

決策樹是由分類與迴歸樹(Classification and Regression Trees, CART)和卡方自動歸納法(Chi-Squared Automatic Induction Detector)等技術產生的強力模型。決策樹應用在監督式資料採礦上，尤其是資料分類，能夠將訓練模組的紀錄區分為獨立的子群，每一子群都有自己的規律。決策樹的優勢在於它的可解釋性高，因為它採用的規則都由直接淺明的文字寫成，這可以讓使用者評價結果、過程中找出關鍵屬性。而在應付不明性質的情況也將特別有用，因為它的直接淺明之規定，可以很容易找出錯誤的結論。同時也能輕易的轉化成邏輯敘述，因此它們可以直接運用到新資料上。本研究採用決策樹方法，經由大量資料分析，可以歸納產品品質問題原因規則，並存於專家系統之中，以完成品質問題診斷與知識擷取的系統。

謝昌倫(2003)歸納決策樹的演算法如下:

- 1.將訓練樣本的資料放入決策樹的樹根。
- 2.將原始資料分成兩組，大約2/3的資料為訓練組，剩餘部分為測試組。
- 3.使用訓練組資料建立決策樹，而在每一個內部節點，評估選擇何屬性可做為分支的依據，如圖3.7所示。
- 4.使用測試組資料進行決策樹修剪，以提升預測速度及避免過適化(overfitting)的問題。
- 5.以上步驟遞迴進行，完成決策樹。以期萃取知識及規則，可用於實際預測或作為參考。

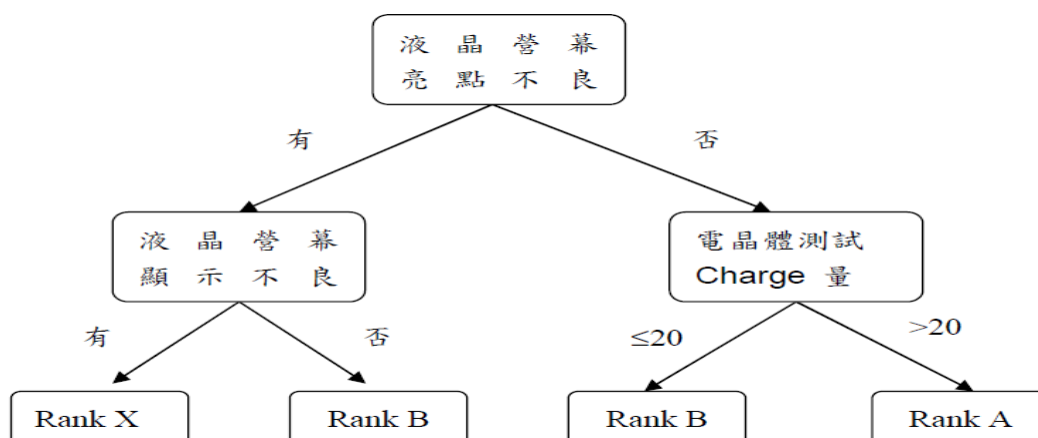


圖 3.7 決策樹分支結構

Jiawei Han & Micheline Kamber(2001)指出現今決策樹的演算法大致有

C5.0、CART、CHAID與QUEST這四種演算法。CART與C4.5/C5.0演算法的最大相異點是其在每一個節點上皆採用二分法，也就是一次只能有兩個子節點，C4.5/5.0則在每一個節點上可以產生不同數量的分枝。CART模型適用於目標變數為連續型和類別型的變數，如果目標變數是類別型變數，則可以使用分類樹(classification trees)，目標變數是連續型的，則可以採用迴歸樹(regression trees)。Perreault和Barksdale兩人於1980年提出CHAID(Chi-square automatic interaction detector)分析方法，CHAID是由AID演變而來的。CHAID會防止資料被過度套用並讓決策樹停止繼續分割，依據的衡量標準是計算節點中類別的P值大小，以此決定決策樹是否繼續分割，所以無需做樹枝修剪。

澳洲學者昆蘭(Qulian)所開發的決策樹演算法ID3(Iterative Dichotomiser 3, C4.5的前身)使用資訊增益(Information Gains)作為分岔準則，說明如下：

1. 資訊理論為若一事件有k種結果,對應的機率為 p_i 。則此事件發生後所得到的資訊量I(視為Entropy)為：

$$I = -(P_1 \times \log_2(p_1) + p_2 \times \log_2(p_2) + \dots + p_k \times \log_2(p_k)) \quad (3.2.11)$$

2. 資訊增益(Information Gain)為若分類標記(Y)分為(成功、失敗)兩種，X為預測變項(類別屬性;k類)，n為總樣本數(n_1 為總樣本數中具成功標記的個數)，經由X變項將樣本分類後 m_{i1} 為X=i類中的總樣本個數(m_{i1} 為X=i類中具成功標記的個數)。根據變項X將n個樣本分為 m_1, m_2, \dots, m_k 的資訊增益為： $Gain(X) = I(n, n_1) - E(X)$ ，其中

$$I = \left(\frac{n_1}{n}\right) \log_2\left(\frac{n_1}{n}\right) + \left(1 - \frac{n_1}{n}\right) \log_2\left(1 - \frac{n_1}{n}\right) \quad (3.2.12)$$

$$E(X) = \left(\frac{m_1}{n}\right) \times I(m_1, m_{11}) + \left(\frac{m_2}{n}\right) \times I(m_2, m_{21}) + \dots + \left(\frac{m_k}{n}\right) \times I(m_k, m_{k1}) \quad (3.2.13)$$

依據ID3演算法，若今天範例資料有三個屬性A、B、C，每個屬性具有兩個可能值，判斷情況有三種，利用ID3可以產生如圖3.8結果(設三個屬性之資訊增益值均大於屬性相關門檻值)。

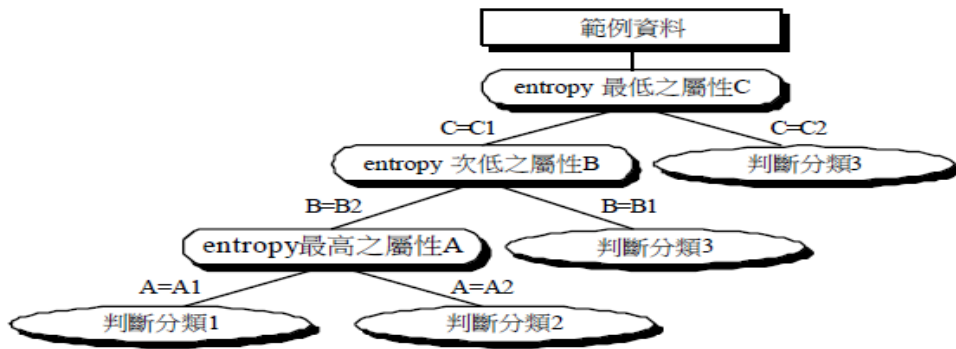


圖 3.8 ID3 決策樹歸納結果示例

本研究所採用的決策樹演算法是由Quinlan於1993年所提出的C4.5演算法，他是改良自ID3學習系統。近年來新發展的歸納學習方法，大多改良自C4.5學習方式及其比較學習效能。C4.5的學習方法包含下列三個步驟:1.由訓練資料建構決策樹 2.決策樹的修剪(prune) 3.從決策樹中產生學習規則，以下將一一說明之:

1.建構決策樹

C4.5學習方法首先就是從訓練資料中建構決策樹，其基本概念可以追朔至1950年代末期的Hoveland及Hunt二位學者所提出的理論。說明如下:假設一組訓練資料所組成的集合S中有K種類別， $S=C_1, C_2, C_3, \dots, C_K$ ，在建構決策樹時有三種情形可能發生:

- (1)當S中的所有訓練資料都屬於同一類別 C_j 時，所建構的決策樹只包含一片樹葉，此片樹葉代表 C_j 的所有資料。
- (2)當S中沒有任何訓練資料時，所建構的決策樹依然只包含一片樹葉，此片樹葉代表的類型則由集合T以外的訓練資料決定。
- (3)當S中包含多種類別的訓練資料時，則將集合T根據某種屬性分割成多種子集合 $S_1, S_2, S_3, \dots, S_n$ ，每個子集合盡可能包含一種類別的資料。S所建構的決策樹包含一個判斷節點及n個分支，每一個子集合的訓練資料分別對應T的一個分支。

一般我們在給定訓練資料時，會有多種決策樹可以正確的將資料分類，因此如何找出一個最小而且能夠正確分類的決策樹便在於分類屬性的考量。本研究採用的C4.5之前身ID3的分類屬性選擇標準稱為gain，他是以測量資訊量多寡來計算各個類別的資訊量，並進而計算出該訓練集

合的平均資訊量，也就是所謂的熵度(Entropy)來表達該集合中資料的複雜度。

假設訓練資料形成的集合S中有n種類別 $C_i = 1, 2, 3 \dots, n$ ，每個類別的資料個數以 $\text{freq} = (C_i, S)$ 表示， $|S|$ 代表S中所有資料的個數，因此各個類別其資料出現機率可表示為 $\frac{\text{freq}(C_i, S)}{|S|}$ ，各個類別的資訊為 $-\log_2\left(\frac{\text{freq}(C_i, S)}{|S|}\right)$ ，訓練集合中包含各個類別的訓練資料，由各類別的資訊量可以計算出訓練集合的平均資訊量(即熵度)，為所有各個類別的資訊量乘上各個類別資料的出線機率總和為：

$$\text{info}(S) = -\sum_{i=1}^n \frac{\text{freq}(C_i, S)}{|S|} \log_2\left(\frac{\text{freq}(C_i, S)}{|S|}\right) \quad (3.2.14)$$

根據 $\text{info}(S)$ 的計算方式，當集合S根據某個屬性A分割成多個子集合 $S_1, S_2, S_3 \dots, S_m$ 時，其分割後所佔的資訊量等於各個子集合的資訊量乘上各個子集合所佔比例的總和：

$$\text{info}_A(S) = -\sum_{i=1}^n \frac{|S_i|}{|S|} \times \text{info}(S_i) \quad (3.2.15)$$

因此集合S經由屬性A分割後所獲得的資訊量則為分割前的資訊量減去分割後的資訊量，表示為 $\text{gain}(A) = \text{info}(S) - \text{info}_A(S)$

而ID3學習系統選擇分類屬性的方法及計算所有屬性的gain值，並選擇其中最大的作為分類屬性。決策樹以此屬性的性質分割成多個訓練子集合。各子樹重複上述步驟，在尚未被選為分類的屬性中再找出gain值最大的作為分類屬性，再分割成多個子樹直到不能再分割為止。

ID3選擇分類屬性的方法對於一般學習問題已有不錯的結果，但是當分類條件較偏向分出的子集合較多的屬性，最特殊的便是當集合S分割後的子集合都只有一個資料時，分割後的資訊量為零，所獲得的資訊量最大，然而此種分割並沒有太大的意義。為了彌補這種缺點，Quinlan在C4.5中提出將gain正規劃的方法以緩和分成過多子集合的效應。

2. 修剪決策樹

在建構決策樹的程序中，有兩個停止的條件，一個是分割後的每個子集合中的資料都屬於同一類別，第二個則是子集合中已經沒有任何資

料可以改進。因此當資料不完整、過於稀疏或是有雜訊時，利用此種方式所建構的決策樹通常過於配合資料(Over-fits the Data)，以至於所產生的決策樹太過複雜，因此再利用C4.5建構決策樹後還需要做適當的修剪。

C4.5修剪過程的標準是根據預估錯誤率(Predicted Error Rate)的值做為判斷條件。他的方法是從樹的底部往上測試每一個節點所形成的子樹，若是將子樹以一個樹葉代替後所得到的錯誤率較低，則將子樹修剪為樹葉，否則將保留原來的子樹，而所謂預估錯誤率是指由現有訓練資料的錯誤比率來評估其他非訓練資料的錯誤比率。更明確的說，假設被分類於某一子樹的所有訓練資料個數為N，其中有E個訓練資料分類錯誤，而錯誤預估率就是由E/N來評估當有新的資料被測試時利用此一條件判斷結果為不正確的機率值。

計算錯誤預估率的方法有數種，C4.5學習系統的預估錯誤率是採用(N,E)二項式分配(Binomial Distribution)的機率值，其做法是給一個可信度(Confidence Level, CF)，簡單的以其二項式分配機率的上限值來做為預估錯誤率的值，寫為UCF(E,N)。其中N是所有訓練資料中屬於此子樹的個數，而E則是N中分配錯誤的資料個數，因此預測錯誤的資料個數則為 $N \times UCF(E,N)$ 。

3. 由決策樹中產生規則

在建立決策樹的分類模型後，我們的直覺是要達到正確的預測，但是除此之外我們還會希望進一步建立類似人類智慧的分類模型。換句話說，分類模型除了要具有高正確率之外最好也能夠擁有簡單易懂的規則。

由決策樹產生規則最簡單的方式便是將決策樹中的每一個樹葉根據其路徑建立一條規則，然而此種方法所產生規則的複雜程度與原來的決策樹一樣，並沒有做任何的改進。而C4.5利用產生規則的過程中做了一次簡化的步驟，以仔細評估每一個條件在過程中的必要性，讓建立規則所使用的條件最精簡，但又不造成錯誤率過高。

C4.5評估規則的方法所使用的評估標準與修剪決策樹的標準一樣是根據錯誤預估率的原則，基本做法如下：

假設有一條規則為：

$$\mathbf{R: If Condition = A Then Class = C} \quad (3.2.16)$$

而另外一個比R更一般化的規則 $\mathbf{R^-}$ 為：

$$R^-: \text{If Condition} = A^- \text{ Then Class} = C^- \quad (3.2.17)$$

條件集合 A^- 為條件集合 A 去除某一條件 X 的結果，寫成 $A^- = A - X$ 。而滿足或不滿足條件 X 在類別 C 與不在類別 C 的資料個數可以表3.1表示：

表3.1 條件 X 屬於分類 C 的對應表

	類別 C	其他類別
滿足條件 X	Y_1	E_1
不滿足條件 X	Y_2	E_2

因此根據規則 R ，會有 $Y_1 + E_1$ 個資料被設定為類別 C ，但其中有 E_1 個資料是屬於其他類別的分類錯誤，因此規則 R 的錯誤預估率為 $UCF(E_1, E_1 + Y_1)$ 。同理，根據規則 R^- ，由於少了條件 X 的判斷，因此會有 $Y_1 + Y_2 + E_1 + E_2$ 個資料被設定為類別 C ，但其中有 $E_1 + E_2$ 個資料屬於其他類別是錯誤的分類，因此規則 R^- 的錯誤預估率為 $UCF(E_1 + E_2, Y_1 + Y_2 + E_1 + E_2)$ 。如果 R^- 的錯誤預估率比 R 的錯誤預估率低，則表示條件 X 不會影響預測錯誤率因此可以被刪除。利用這個觀念，C4.5的學習系統為求計算方便，改於在每個規則中刪除最沒有貢獻的條件。

對於決策樹分類的評估方法上，Kusiak & Kurasek(2001)使用十摺交叉驗證模式(10-fold)來建構交叉驗證架構，十摺交叉驗證方法為分析資料分成十組(subset)，每一組按順序與其他九組組成一摺資料組，在每摺的資料組中，包含了有九份訓練資料組及一份測試資料組，將每摺之測試資料組經由決策樹分類規則產生預測結果，如此反覆進行十次，再統計十摺之平均失誤率作為評量指標。本研究使用交叉確認架構來進行決策樹評估。研究最終，將所有蒐集的資料利用該架構進行分析，以進行分類結果效度評估。

第四章 實證分析

4.1 問題定義與資料屬性定義

4.1.1 問題定義

本研究將問題定義在品質問題的分類分析上，製品的品質問題可能發生在製程之中，也有可能因為漏檢或者是先前製程的錯誤，導致後續製品的不良，最嚴重的是這些不良品最終落到顧客手上，而造成商譽或者成本上的損失。若今天顧客針對產品提出問題，則必須快速找到以往顧客問題資料是否有相似的問題，同時也必須快速反映至製造現場，並從與問題有關之資料中，找到造成問題的因素。因此，本研究根據問題及目的，蒐集相關的品質問題處理資料，並選擇合適的方法進行分析，對於品質問題原因考慮因素而言，考慮到控制製程的人員因素、機器設備因素、材料因素、環境因素、方法因素等。建立品質問題診斷最主要是希望一旦有品質問題發生時，資訊能迅速呈現，以利於品質負責人員能正確根據分類推論的知識，快速做出適當的改正措施。

4.1.2 資料準備與屬性定義

本研究的資料來源為一家本土中小尺寸TFT-LCD製造廠商的品管資料，其中是以CELL的資料為主要分析對象，資料的蒐集方法為實際的訪談，透過與該公司相關人員的面對面交談，一方面了解該公司的製造過程及品質異常資料蒐集方式，一方面解釋本研究之目的與所需要之資料類型，經過訪談所獲取的資料其蒐集時間為民國95年至民國97年，個案公司在CELL製程中的不良件報告，這樣的資料量就資料採礦方法來說是有些許不足，但沒有更往前做進一步分析的原因是因為這家TFT-LCD製造廠商是在近年才向另一家大尺寸TFT-LCD製造廠商收購了這座生產工廠，舊有資料已經隨著廠區的移轉而銷毀或是報廢，因此造成資料蒐集不易。資料經過蒐集與彙整後的筆數總共為215筆，由於不良成因可能為一個或多個，必須資料清潔與整理之後才能進行分析，最後總共彙整出計有36種品質問題類型。

現今大多數的生產工廠對於製程與品管資料的呈現都是以表單紀錄並尋找品質問題發生原因，本研究的資料來源也是以表單型式呈現，進而對

表單理的資料進行資料庫的建立，但是為了要確保分析上的準確性，必須先將資料做轉化編碼處理，而不是把所有資料一次全部放進分析模式。本研究之個案公司在進行不良件報告檢討時，以內容來說主要可分為六大項：

- 1.不良現象:即品管人員在檢測的過程中所發生不良的情形，如輝點、線不良、配向膜內異物等。
- 2.點燈現象:為TFT-LCD的測試過程中重要的步驟，利用檢測儀器可看出確實的不良情形，如畫素呈現亮點、縱向或橫向成一線狀輝點、畫面下為白色等。
- 3.可能成因:品管人員對於不良發生的原因作各種原因的探究，針對人員、機器、材料、方法與環境等細項作進一步分析，如TFT元件異常、靜電破壞、彩色濾光片畫素異常等。
- 4.解析方法:利用更精密的儀器或更進一步的動作去檢測不良發生成因，例如將不良品拆解或是於顯微鏡下進行觀測動作，如TFT去掉配向膜,直接於顯微鏡觀察、點燈確認低頻閃爍、端子部顯微鏡觀察等。
- 5.資料查詢或對策:查詢是否為料件不良或是機器問題，並確認各生產過程有無疏失，再與之前發生過類似原因作比較，來擬定日後的改進對策，如靜電產生機構查詢、根據刮傷方向及膜上膜下判定可能機構,機台是否調整?面板是否為再生品?
- 6.對策單位:明確指出應該要對此缺失負責並改進的單位，如生產、工程、ARRAY等。

本研究採用之個案公司資料中的不良件報告，在進行產品缺陷報告時，使用的是五個變數進行分析:即作業人員(manpower)、機器(machinery)、材料(material)、作業方法(method)及作業環境(environment)(鄭春生,1995)，而不良件中缺陷的來源可能是以上五個成因中的一個或多個，假若以不良件中的「線不良」此原因作分析，則可利用以下的特性要因圖表示之:

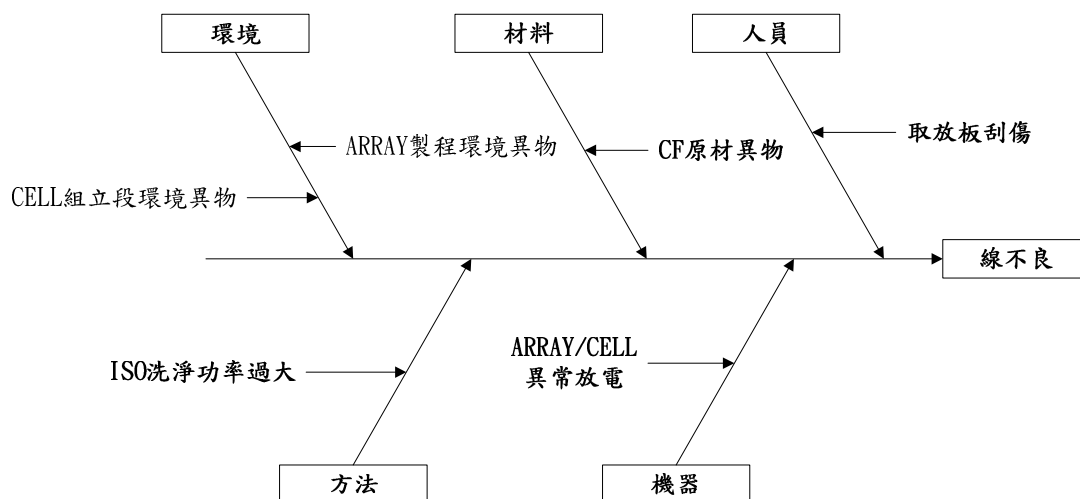


圖 4.1 造成線不良之特性要因圖

(資料來源:本研究整理)

特性要因圖中分析的原因，為品管人員一一的詢問各個製程的工程師在每個製程中，有哪些因素會造成產品的缺陷。本研究主要是結合各製程工程人員的知識先匯整出所有可能造成品質異常的原因類型，再與蒐集到的資料所分析出來的結果作比較，以往檢測人員在作品質問題檢驗時往往需要用人工目視的方法作全檢的動作，透過本研究的分析歸納後，當品管人員下次遇到同樣的問題時，可依分類出的知識，快速的作出生產方面的改進決策，如此或許可省去每一次在發生品質異常的情形時，都需要作一次全檢的動作的時間，而提升整體的效率。

進行分類歸納的資料記錄中，主要欄位必須包括貨批編號、屬性以及判斷。經由事先對於品質問題資料的檢視，大致對於資料屬性的定義成五個屬性，主要分為人員、設備、參數、方法、材料等五個屬性 (即4M+參數)，在數個屬性下所對應的資料分別記載「無錯誤發生」或「發生錯誤狀況」。對於資料格式的表示，通常可分為名義變數、次序變數、區間變數、比率變數等，在資料格式的輸入，在結果解釋上來說，名義變數最為容易作為解釋的格式。決定完屬性及其對應可能值後，需對判斷欄的格式做定義，判斷欄通常為品管人員檢驗出來的品質問題症狀，並且在進行資料輸入前，則需於事先對品質問題處理資料上的各不良品報告做整理，例如對於不良發生情形的描述需統一。資料整理類型如表4.1所示：

表4.1 參數編碼示例

不良編號	人員	機器	材料	方法	環境	品質問題
1	人員類型1	機器類型1	材料類型1	方法類型1	環境類型1	品質問題1
2	人員類型2	機器類型1	材料類型3	方法類型2	環境類型1	品質問題2
3	人員類型1	機器類型1	材料類型3	方法類型3	環境類型1	品質問題3
:	:	:	:	:	:	:
215	人員類型1	機器類型1	材料類型1	方法類型0	環境類型1	品質問題1

(資料來源:本研究整理)

在資料的編碼過程中，若一項不良品其造成的原因並不包含人員、機器、材料、方法、環境此五項可能成因中的一項或多項，則我們將不包含的項目編碼定義為「0」，表示其只有四項以下之因子影響到最後的成因。而每次相同的品質檢測不良結果其背後可能造成之原因也有可能不同，例如品質問題1在上表中是由人員原因1、機器類型1、材料類型1、方法類型1及環境類型1所造成，但是也會出現此五項中有一項正常的情況下，仍然出現品質問題1的情形，此時我們就把這項品質問題視為兩種不同的分析項目。依循此原則下，我們共歸納出10種人員原因變數、11種機器原因變數、12種材料原因變數、14種方法原因變數及7種環境原因變數，如表4.2~4.6這些原因包含在本研究所蒐集的215條不良件報告中，並彙整出36種品質問題類型。

表4.2 人員原因變數

編號	人員原因變數	編號	人員原因變數
1	靜電裝置未正確配戴使用	6	取放板刮傷面板走線
2	碰觸污染膜面	7	轉寫刮刀裝配不平整
3	配向布裝配不平整	8	轉寫 APR 版未洗淨
4	取用液晶污染	9	刮膠清潔不全
5	熱壓固著前手動取放板	10	按壓面板力道過大

(資料來源:本研究整理)

表4.3 機器原因變數

編號	機器原因變數	編號	機器原因變數
1	ARRAY/CELL 機台除靜電設備異常或異常放電	2	PI 轉印品質不良
3	洗淨毛刷髒污	4	設備干涉刮傷 PI 膜
5	配向機震動	6	機台 stage 不平整，吸著壓過大
7	框膠塗佈機 / 注入機 / 加壓封止機異常	8	組立移載手臂偏移
9	散佈機靜電異常	10	偏貼機構干涉發塵
11	液晶皿污染		

(資料來源:本研究整理)

表4.4 材料原因變數

編號	材料原因變數	編號	材料原因變數
1	CF 原材異物	2	spacer 受潮
3	TFT 來料不良	4	刮刀刃不平整
5	框膠原材/異物	6	機板排版設計不易排氣
7	配向 Rollor 真圓度不足	8	液晶原料污染
9	面板 Short Ring 設計無法有效導掉靜電	10	偏光板原材異物
11	UV 封止膠過期或污染	12	MDL 材搭配性

(資料來源:本研究整理)

表4.5 方法原因變數

編號	方法原因變數	編號	方法原因變數
1	前洗壓入量不足	2	PI 吐出量設定
3	轉寫速率設定	4	配向參數設定
5	框膠爐溫參數設定	6	注入機參數設定
7	加壓封止機參數設定	8	ISO 洗淨功率參數設定過大
9	熱壓組立偏移參數設定	10	ARRAY 成膜參數
11	偏貼速度參數設定	12	剝膜機參數設定
13	切裂參數設定	14	點燈參數設定

(資料來源:本研究整理)

表4.6 環境原因變數

編號	環境原因變數	編號	環境原因變數
1	ARRAY 製程環境異物	2	CELL 組立段環境異物
3	PI 預烤部風量不穩	4	框膠爐區風量過大
5	spacer 保存環境溼度	6	溼度過低
7	偏貼段環境異物		

(資料來源:本研究整理)

4.2 決定分群數目

4.2.1 華德法初步分群

本研究使用 spss17.0 統計分析軟體來對此進行分析，首先進行二階段分群法的第一階段，也就是階層式分群法中的華德法(Ward method)，此方法主要是用來確定資料適合的分群數目，計算區間採用歐基里德直線距離平方(Squared Euclidean distance)來觀察 215 筆資料由 215 群凝集至 1 群的凝集過程，結果如表 4.7 所示。

表4.7 群數凝聚過程

階段	組合集群		係數	先出現的階段集群		下一階段
	集群1	集群2		集群1	集群2	
1	197	214	.000	0	0	141
2	109	212	.000	0	0	69
3	45	210	.000	0	0	185
4	44	209	.000	0	0	196
:	:	:	:	:	:	:
:	:	:	:	:	:	:
210	1	7	3468.265	206	190	213
211	5	6	4130.195	209	208	212
212	5	11	4901.319	211	201	214
213	1	30	5877.185	210	207	214
214	1	5	8839.581	213	212	0

(資料來源:本研究整理)

從資料表中可以看出，全部資料從215群凝集至1群的過程中，其變異

係數(coefficient)在二群要凝集為一群時差異最大，因為在係數值愈小時表示該小組成員愈同質，反之係數值愈大表示該組成員相異性最高，因此我們可把本研究的資料分成兩群，但因為假使分群數太少，會導致最後在結果判讀時發生判斷錯誤的機會較大，因此兩群的資料亦可分成五群(張紹勳等,2000)。

本方法在判斷分群數時所使用區集為歐基里德直線距離平方，其意義是希望每一分群裡的數值與其中心點的距離要愈近愈好，因此接著比較分為兩群與分為五群與中心距離平均數之差異。

表4.8 二集群因子平均數

Ward Method		manpower	machine	material	method	environment
1	平均數	3.6944	3.3194	3.6944	1.5139	2.0069
	個數	144	144	144	144	144
	標準差	2.21968	2.44850	2.31226	1.58660	1.53512
2	平均數	1.8028	4.9155	3.2113	8.9577	1.2817
	個數	71	71	71	71	71
	標準差	2.95016	3.29825	3.03275	2.90043	1.89121
總和	平均數	3.0698	3.8465	3.5349	3.9721	1.7674
	個數	215	215	215	215	215
	標準差	2.63332	2.85141	2.57549	4.09230	1.69161

(資料來源:本研究整理)

表4.9 五集群因子平均數

Ward Method		manpower	machine	material	method	environment
1	平均數	3.2778	2.7540	3.0317	1.3254	2.0000
	個數	126	126	126	126	126
	標準差	1.97034	1.85337	1.48559	1.41325	1.27122
2	平均數	4.3103	3.4138	1.3103	9.4138	1.0690
	個數	29	29	29	29	29
	標準差	3.26347	3.01801	1.00369	1.99136	.88362

Ward Method	manpower	machine	material	method	environment
3 平均數	.0000	7.4848	3.2727	7.4545	1.2727
個數	33	33	33	33	33
標準差	.00000	1.00378	2.03520	2.69364	2.16900
4 平均數	.3333	.3333	9.1111	13.0000	2.0000
個數	9	9	9	9	9
標準差	.50000	.50000	2.84800	1.50000	3.00000
5 平均數	6.6111	7.2778	8.3333	2.8333	2.0556
個數	18	18	18	18	18
標準差	1.61387	2.49247	1.64496	2.09341	2.81743
總平均數	3.0698	3.8465	3.5349	3.9721	1.7674
總個數	215	215	215	215	215
標準差	2.63332	2.85141	2.57549	4.09230	1.69161

(資料來源:本研究整理)

比較分成五集群與分成二集群之間第一群之平均數差異，可以看出在人員成因(五集群3.2778;二集群3.6944)、機器成因(五集群2.7540;二集群3.3194)、材料成因(五集群3.0317;二集群3.6944)、方法成因(五集群1.3254;二集群1.5139)、環境成因(五集群2.0000;二集群2.0069)等五個影響因素皆是分成五集群的情況之下，其平均數與中心點之距離較二集群來的近，代表分成五集群之分群效度比分成二群要來的好，而雖然在五集群中之第二、三、四、五集群之分群效度並沒有全部都比分成二集群為佳，但是由於這些集群也是由原本的二集群所分裂，加上也有部分比原本的分群為佳，例如五集群中的機器、材料、環境三個成因都比二集群為佳，因此可以斷言，此資料分為五個集群是適當的。

4.2.2 K平均數法分群

在初步把資料分為五群之後，利用K-means方法對於資料作一些觀察與驗證的動作。以第一階段分群所得到的結果五群做為分群數，剛開始K-Means Cluster會隨機指定各個群組中心，並且進行計算初始群集的中心點，如表4.10，接著計算每一筆商品的八個輸入向量到各集群重心的歐基

里德距離，隨著計算疊代的變化，各集群的商品向量會向中心點收斂而改變群組的個數。

表4.10 初始集群中心點

	集群				
	1	2	3	4	5
manpower	3.28	4.31	.00	.33	6.61
machine	2.75	3.41	7.48	.33	7.28
material	3.03	1.31	3.27	9.11	8.33
method	1.33	9.41	7.45	13.00	2.83
environment	2.00	1.07	1.27	2.00	2.06

(資料來源:本研究整理)

本研究的不良品成因紀錄一共為215筆，在經由3個疊代的演算之後，叢集中點沒有或僅有小幅變動因此達成收斂，如表4.11每一個集群中心點的商品向量的最大距離已收斂為0，如圖4.2，並且得到最後集群中心點商品的五個輸入向量，如表4.12，因此結束演算疊代，並且確定每個集群的數目。

表4.11 疊代記錄^a

集群中心點的變更	疊代		
	1	2	3
1	.152	.105	.000
2	.291	.000	.000
3	7.684E-5	.000	.000
4	4.843E-5	.000	.000
5	.863	.488	.000

(資料來源:本研究整理)

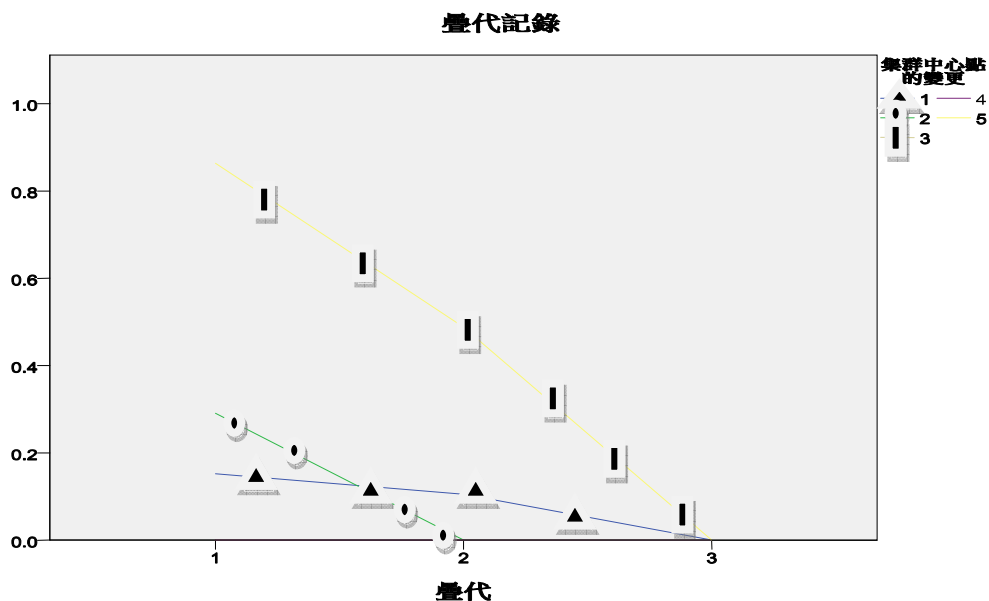


圖 4.2 K-means 疊代收斂圖
(資料來源:本研究整理)

表4.12 最後集群中心點

	集群				
	1	2	3	4	5
manpower	3.08	4.14	.00	.33	7.00
machine	2.62	3.25	7.48	.33	7.04
material	3.01	1.14	3.27	9.11	7.25
method	1.38	9.43	7.45	13.00	2.50
environment	2.08	1.11	1.27	2.00	1.54

(資料來源:本研究整理)

在K-means Cluster分群模式下，將所有的不良品記錄分為五個群組，第一個群組的不良品數為121個，第二群組的不良品數為28個，第三個群組的不良品數為33個，第四個群組的不良品數為9個，第五個群組的不良品數為24個，總資料筆數為215筆，如表4.13所示，要注意的是，K-means方法所分出來的群組在本研究中並不就是當作最後的分析數據，還要與SOM分群出來的結果做比較。

表4.13 各集群中的觀察值個數

1	121.000
2	28.000
3	33.000
4	9.000
5	24.000
有效的	215.000
遺漏值	1.000

(資料來源:本研究整理)

4.2.3 對K平均數法結果做驗證

在研究中做分群動作的目的，是希望能使分群後的結果呈現組間差異大，組內差異小的情形，假使最後資料能顯現這種形式，則代表分群結果非常具有效力與解釋力，反之則是無效的分群，現將本研究所採用的資料之結果以下表4.14做說明。

表4.14 分群結果ANOVA表

	集群		誤差		F 檢定	顯著性
	平均平方和	自由度	平均平方和	自由度		
manpower	195.338	4	3.346	210	58.384	.000
machine	246.243	4	3.595	210	68.495	.000
material	201.783	4	2.916	210	69.199	.000
method	708.070	4	3.579	210	197.850	.000
environment	8.504	4	2.754	210	3.088	.017

(資料來源:本研究整理)

在ANOVA表中，能夠代表集群之間差異的是集群的平均平方和(cluster mean square)，而能代表組內差異的是誤差的平均平方和(error mean square)，兩者分別除以其自由度(degree of freedom)的比值即是F值，故從上表可以看出，若我們欲將本研究之215筆品質不良資料分門別類，則人員、機器、材料、方法等四種因素因其F值均夠大，且P值皆為 $0.000 < 0.05$ ，由此著手都可有效的將資料分類，而環境因素雖其F值較小，但因為P值為0.017

也小於0.05，因此也可考慮將此因素納入分類的規則當中。

為了驗證分群之效益，因此繼續對結果做判定分析，判定分析與集群分析相同，都是將相似的事物歸為一類，不同處在於集群分析預先不知道分類，而判別分析是在研究對象分類已知的情況下，根據樣本資料推導出一個或一組判別函數，同時指定一種判別規則，用以確定待判別樣本所屬的類型，使錯判率最小。由表4.15得知，Box's M共變異數相等性檢定的近似值為11.105，其顯著性為 $0.000 < 0.05$ ，因此具有顯著水準，符合進行判定分析的條件。

表4.15 檢定結果^a

Box's M 共變數相等性檢定	236.461
F檢定 近似值	11.105
分子自由度	20
分母自由度	15602.582
顯著性	.000

(資料來源:本研究整理)

表4.16 Wilks' Lambda值

函數檢定	Wilks' Lambda值	卡方	自由度	顯著性
1 到 4	.016	871.682	16	.000
2 到 4	.084	519.893	9	.000
3 到 4	.260	282.292	4	.000
4	.632	95.989	1	.000

(資料來源:本研究整理)

表4.16中之Wilk's Lambda值即代表總變異數中可由組內變異數來解釋的比例。Wilk's值愈大，表示組平均數愈沒有差異，當此值等於一時表示所有觀測組的平均數皆相等，此值愈小，表示組的平均數差異愈大，此值接近零時，表示總變異數中大多是受到組間平均數差異的影響(張紹勳等,2000)。因為本研究已將組別分為五組，因此判別函數只要四個，即從四個維度進行觀察值分類(林震岩,2007)，此表為向度縮減分析，亦即在考驗

區別函數的顯著性(吳明隆&涂金堂,2005)，首先先考驗所有的區別函數，如表中Wilk's Lambda值=0.16，卡方值871.682， $P=0.000<0.05$ 已達顯著水準，表示四個區別函數中，第一個區別函數對依變項有顯著的解釋力，其次，剔除第一區別函數的影響後，第二區別函數的Wilk's Lambda值=0.84，卡方值519.893， $P=0.000<0.05$ 也達到顯著水準，表示第二個區別函數對依變項的解釋力也達顯著，依此規則類推，本研究在第三及第四區別函數對依變項的解釋力也達到顯著。

表4.17 標準化的典型區別函數係數

	函數			
	1	2	3	4
manpower	-.335	.896	.130	.637
machine	.187	-.993	.481	.337
material	.287	.423	.789	-.423
method	.949	.363	-.149	.154

(資料來源:本研究整理)

表4.17為四個典型區別函數的標準化係數，稱為標準化的典型區別函數係數(Standardized Canonical Discriminant Function Coefficients)，標準化區別係數代表各自變項在各區別函數上的相對重要性，係數值愈大，代表該自變項在區別函數的重要性愈大。從表中可看出，在第一個區別函數中，方法變項與其關係較密切，也就是重要性愈大，第二區別函數為人員變項，第三區別函數為材料變項，第四區別函數為人員變項，若與SOM方法比較後K-means的分群效能較佳，則此結果將用來作為決策樹分群依據之用，如表4.18。

表4.18 各組重心的函數

集群觀察 值個數	函數			
	1	2	3	4
1	-1.512	.021	-.350	-.294
2	2.080	1.291	-1.610	1.305
3	2.625	-2.805	.310	-.073
4	5.620	3.612	.780	-2.223
5	-.520	.890	2.923	.894

(資料來源:本研究整理)

表4.18為在進行決策樹分類時，應以何項作為分類依據的準則，如在第一群中，只有在第二區別函數時為正值且為最大的數字，因此在進行第一群決策樹分類時，應以人員變項為主要的分類依據，同理第二群應以方法變項為分類依據，第三群應以材料變項為分類依據，第四群應以方法變項為分類依據，第五群應以材料變項為分類依據。

4.3分群結果

4.3.1 K-means分群結果

在上一節的討論過後，已經確定把資料以K-means方法分成五群的結果是可行的，因此以下將把各群資料以表列方式呈現並加以說明：

表4.19 K-means第一群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
配向膜異物(3)	51,52,128,129,130,148,187,199,200	9	7.4%
面板內異物(4)	2,53,107,131	4	3.3%
縱不均(5)	1,27,28,136,138,139,140,155,156,169,170,171,176,177,178,179,180,203	18	14.9%
橫條紋不均(6)	3,13,29,41,42,43,62,63,64,65,153	11	9.1%
T不均(7)	4,9,12,49,50,71,72,85,86,87,88,104,135,142,151,154,172,186,188,197,202,206,211,214	24	19.8%
配向條紋A(8)	95,105	2	1.7%
配向縱條紋(10)	8,	1	0.8%

問題類別	紀錄編號	合計筆數	所佔比率
黑點(11)	149,150,160,161, 174,175	6	5%
白點(12)	164	1	0.8%
線狀微輝點(13)	165	1	0.8%
半月GAP(15)	79,115	2	1.7%
面內GAP(17)	21	1	0.8%
膜面刮傷(21)	55,56,59,60,61,69,70,81,106,117,118,1 52,183,184,208	15	12.4%
SHOT(28)	7,31,103,192	4	3.3%
雨狀不均(31)	93,127,133,	3	2.5%
注入口附近黑色 區塊(32)	24,94,109,134,212	5	4.1%
面板四角發生 GAP(33)	16,25,110,141,213	5	4.1%
CSOP(34)	26,111,137	3	2.5%
1/3橫條紋(35)	82,97	2	1.7%
PKISS(36)	37,54,77	3	2.5%

(資料來源:本研究整理)

表4.20 K-means第二群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
輝點(1)	5,124,145,166,173,189,195	7	25%
線不良(2)	123,146,147,167,168,190,198	7	25%
T不均(7)	22	1	3.6%
面內GAP(17)	120,205	2	7.1%
異物GAP(18)	121,196	2	7.1%
CF裂化(22)	44,209	2	7.1%
碰傷(23)	45,46,73,210	4	14.3%
漏光(25)	35,48,75	3	10.7%

(資料來源:本研究整理)

表4.21 K-means第三群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
配向條紋B(9)	14	1	3%
周邊GAP(16)	17,18,19,20,36,38,39,40,80,116, 119,204	12	36.4%
異物GAP(18)	23	1	3%
SPACER打痕或凝集(19)	6,10,112,113,122,181,185,207	8	24.2%
斜視SHOT(29)	32,89,193	3	9.1%
綠不均(30)	90,91,92,125,126,162,163,194	8	24.2%

(資料來源:本研究整理)

表4.22 K-means第四群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
靜電氣(20)	11,68,182	3	33.3%
殘像(26)	100,101,132,157,158,215	6	66.7%

(資料來源:本研究整理)

表4.23 K-means第五群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
配向條紋B(9)	57,66,96,143	4	16.7%
配向縱條紋(10)	58,67,144,159	4	16.7%
角殘(14)	33,34,78,114	4	16.7%
偏光板不良(24)	24,74,201	3	12.5%
漏光(25)	76	1	4.2%
注入口偏白(27)	30,102,191	3	12.5%
PKISS(36)	15,83,84,98,99	5	20.8%

(資料來源:本研究整理)

在這些分群結果當中可以發現，很明顯有分群結果不均的情形，第一群佔了大部分資料數，可能因為K-means演算法的主要缺點為對雜訊(noise)敏感(Jiang et al., 2004)，由於本研究資料的特性為資料維度高，且極可能有相當多的雜訊資料或遺漏值(missing value)，使得階層式分群方法與K-means方法的應用受到了相當大的限制，且本研究主要是希望能夠透過分

群之後的結果，利用決策樹得到分類的規則，但在得出規則之前試著對這些資料做決策樹分析，發現在本研究中的K-means分群方法中的第四群，並沒有辦法得出決策規則，也就是無法得到決策樹的圖形，因此本研究判定K-means分群方法並不適用於本筆資料，因而將以類神經網路的自我組織映射網路繼續分析。

4.3.2 自我組織映射圖分群結果

在前面的文獻探討中有討論過，由於本研究所蒐集的品質不良資料是由品管人員作人工目視全檢所整理出來的，在人力有限的情形之下，難免會有錯誤疏漏之處，而在資料可能有誤的前提下，若使用自組織映射圖網路(SOM)所得到的分群結果會比使用傳統的統計分群方法(本研究為K-means法)要來的正確，且在上一節中指出本研究若利用K-means的分群方法來做分析會有得不出決策樹的情形發生，因此本研究將繼續用SOM方法針對品質不良資料做分群動作。

1. 參數設定

本研究使用MATLAB 7.0應用軟體提供之Neural Network Toolbox模組進行自組織映射圖網路的數據分析，在開始分群之前，要先對參數進行設定，分述如下表：

表4.24 SOM的學習參數設定

學習參數	設定值	描述
Dimensions of map	[5,1]	根據二階段分群結果決定把群數分為五群，因為五已是質數，沒有其他兩兩相乘之數可為五，因此設定值為此型態
Topology function	GRIDTOP	創建一個矩形鄰域(內定值)
Distance function	DIST	此函數代表歐基里德(Euclidean)距離函數
Ordering phase learning rate	0.9	排序階段的學習速率(內定值)
Ordering phase steps	1000	排序階段的步階(內定值)
Tuning phase learning rate	0.02	調諧階段的學習速率(內定值)
Neighborhood distance	1	調諧階段的鄰域距離(內定值)

學習參數	設定值	描述
epochs	1000	設定訓練的最大循環次數(內定值)
goal	0	設定性能目標(內定值)
show	25	設定每隔多少訓練循環數目就顯示 (內定值)
time	Inf	設定最大的訓練所需時間，單位為秒 (內定值)

(資料來源:羅華強(2005)，本研究整理)

其中排序階段(Ordering phase)為在階段中持續給定步階數目，鄰域距離開始時是如同兩個神經元之間最大的距離，而降低到調諧的鄰域距離。學習速率開始時是在排序階段的學習速率，而降低直到達到調諧階段的學習速率為止。隨著鄰域距離和學習速率的降低，此網路的神經原典型上將他們自己排序在輸入空間中並具有相同的拓樸，此處他們將會根據自然法則來排序。

而調諧階段(Tuning phase)持續於訓練或適應的其餘步階內，鄰域距離停留在調諧的鄰域距離(典型上是1.0)，學習速率繼續從調諧階段的學習速率降低，但卻是很慢的進行。小的鄰域和緩慢的學習速率微調此網路，而保持在先前階段內所學習到的排序穩定。針對調諧部分的循環訓練次數，應該是要更加大於在排序階段內的步階數目，因為調諧階段通常會更加耗時。一旦鄰域大小是1，此網路應該被排序的相當好，並且學習速率在一般較長的期間上慢慢降低，以使神經元有足夠時間展開。

2.分群結果

在事先預做的決策樹分析中，得到以SOM方法來分群，每一群都可以順利的分出各自規則的結果，以下將以表格展現本研究以SOM為分群方法得到之分群結果，並對分群內容加以說明：

表4.25 SOM第一群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
配向膜異物(3)	51,128,129,148,187,199,200	7	14.3%
面板內異物(4)	2,53,131	3	6.1%
縱不均(5)	1,136,139,140,155,170	6	12.2%
橫條紋不均(6)	3,41,42	3	6.1%

問題類別	紀錄編號	合計筆數	所佔比率
T不均(7)	4,49,88	3	6.1%
配向條紋A(8)	95,105	2	4.1%
半月GAP(15)	79,115	2	4.1%
SHOT(28)	7,31,103,192	4	8.2%
兩狀不均(31)	93,127,133	3	6.1%
注入口附近黑色 區塊(32)	24,94,109,134,212	5	10.2%
面板四角發生 GAP(33)	16,25,110,141,213	5	10.2%
CSOP(34)	26,111,137	3	6.1%
1/3橫條紋(35)	82,97	2	4.1%
PKISS(36)	37	1	2%

(資料來源:本研究整理)

表4.26為經SOM分群之後第一群之不良品資料，共49筆資料。結果為本群最多的不良類型有兩種，分別為異物的產生及面板條紋發生不均的情形，而次多的不良類型也有兩種，分別為發生GAP的情形及注入口附近產生黑色區塊，查詢不良件紀錄發現原因多半是因為配向膜製程造成，而配向膜製程會有不良發生主要跟人員、機器變項有關，與方法、環境變項有次相關，而材料變項影響的因素最小，因此第一群的結果主要原因出在人員與機器製程的偏誤造成不良品的發生，因此本研究認定人員與機器變項為本群在決策樹分析中的主要判別因子，若與K-means分群方法比較，同樣都是人員變項的影響最大。

表4.26 SOM第二群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
配向膜異物(3)	52,130,	2	3.2%
面板內異物(4)	107	1	1.6%
縱不均(5)	27,28,138,156,169,171,176,177, 178,179,203	11	17.7%
T不均(7)	9,12,50,71,85,87,135,142,151, 154,172,186,188	13	21%

問題類別	紀錄編號	合計筆數	所佔比率
橫條紋不均(6)	13,29,43,62,63,64,65,153	8	12.9%
配向縱條紋(10)	8	1	1.6%
黑點(11)	108,149,150,160,161,174,175	7	11.3%
白點(12)	164	1	1.6%
線狀微輝點(13)	165	1	1.6%
膜面刮傷(21)	55,56,59,60,61,69,70,81,106,117,118, 152,183,184,208	15	24.2%
PKISS(36)	54,77	2	3.2%

(資料來源:本研究整理)

表4.27為經過SOM分群之後第二群之不良品資料，共有62筆，在第二群中以面板條紋不均及膜面刮傷為最多的不良情形，黑點的產生為次多的不良情形，主要原因除了配向膜參數及吐出量設定等方法品質問題之外，此群產生不良的原因還有屬於人員變項相關的刮刀刀不平整及材料變項的彩色濾光片來料不良，而次要原因有屬於環境變項的配向膜預烤部風量不穩，因此本群主要不良成因為人員、方法及材料變項的偏誤，次要原因為環境變項的偏誤，因此本研究將人員、方法及材料認定為本群在做決策樹分析時的主要判別因子，與K-means分群相比較，可以看出判斷的主要肇因變項是一致的。

表4.27 SOM第三群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
縱不均(5)	180	1	3%
T不均(7)	72,86,104,197,202,206,211	7	21.2%
配向條紋B(9)	57,66,96,143	4	12.1%
配向縱條紋(10)	58,67,144,159	4	12.1%
角殘(14)	33,34,78,114	4	12.1%
周邊GAP(16)	21	1	3%
偏光板不良(24)	47,74,201	3	9%
面板四角發生 GAP(33)	214	1	3%

問題類別	紀錄編號	合計筆數	所佔比率
注入口偏白(27)	30,102,191	3	9%
PKISS(36)	15,83,84,98,99	5	15.1%

(資料來源:本研究整理)

表4.28為SOM分群之第三群不良資料，共33筆，分布最多的不良資料為面板不均現象及配向膜上出現條紋，而次多的不良情形為角殘及PKISS，在不良件紀錄中，此群最大影響原因為機器變項的配向機震動及機台上夾管線角度異常或磨損等，而次要影響原因為環境變項的配向膜預烤部風量不穩及框膠爐區風量過大等，因此本群主要影響原因為機器變項，次要影響原因為環境變項，因此本研究將機器與環境變項認定為本群在做決策樹分析時的主要判別因子，與K-means之結果相比較，明顯的與其判別材料變項為主要不良原因有所不同。

表4.28 SOM第四群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
配向條紋B(9)	14	1	2.7%
周邊GAP(16)	17,18,19,20,36,38,39,40,80,116,119,204	12	32.4%
面內GAP(17)	22,120,205	3	11.1%
異物GAP(18)	23,121,196	3	11.1%
SPACER打痕或凝集(19)	10,113,181	3	11.1%
漏光(25)	35,48,75,76	4	10.8%
斜視SHOT(29)	32,89,193	3	11.1%
綠不均(30)	90,91,92,125,126,162,163,194	8	21.6%

(資料來源:本研究整理)

表4.29為SOM分群後之第四群資料，共37筆，由表中可看出此群最大的不良問題在於GAP的產生，尤其是周邊GAP，另外綠不均的問題也在本群中佔多數。在不良件資料的檢視過程中發現機器中的機台配置不平整，吸著壓過大、組立移載手臂偏移及框膠塗佈機、注入機、加壓封止機異常等因素佔此群的大多數不良成因，而次要因素則為材料變項中的SPACER受潮等情形，因此判斷此群主要影響變項為機器變項，次要影響變項為材

料變項，因此本研究將機器及材料變項當作本群做決策樹分析時的主要判別因子。而與K-means的紀錄相比較，明顯的與其判別方法變項為主要不良成因有不同。

表4.29 SOM第五群之品質類型及紀錄編號

問題類別	紀錄編號	合計筆數	所佔比率
輝點(1)	5,124,145,166,173,189,195	7	20.6
線不良(2)	123,146,147,167,168,190,198	7	20.6
配向條紋B(9)	6	1	3
SPACER打痕或凝集(19)	112,122,185,207	4	11.8
靜電氣(20)	11,68,182	3	8.8
CF裂化(22)	44,209	2	5.9
碰傷(23)	45,46,73,210	4	11.8
殘像(26)	100,101,132,157,158,215	6	17.6

(資料來源:本研究整理)

表4.30為SOM分群所分出來的第五群，共有34筆，由表中可看出此群最大的不良問題為輝點、線不良及殘像，在不良件資料的檢視中發現此群最大的不良成因為方法變項中的切裂參數設定、ISO洗淨參數設定及ARRAY成膜參數設定等，而次要影響因素為機器變項中的ARRAY/CELL機台異常或有異常放電情形及散佈機靜電異常等，由此可判定本群主要不良變項為方法變項，次要不良變項為機器變項，因此本研究將方法與機器變項認定為本群做決策樹分析時的主要判別因子。與K-means的分群做比較，明顯的與其判定的材料變項為主要不良因素有差異。

經由SOM分群之後可發現，在分析不良品資料時，以機器變項為主要判別因子最有可能判別出不良成因，其次為人員變項、方法變項及環境變項，而以材料變項為主要判別因子則最不容易判別出不良成因。

4.4 決策樹分析

在上一節中，本研究將215筆資料依據特性的不同予以分群，希望能盡量將資料中的雜訊排除，在本節中，將利用分群後的五群資料分別進行決策樹分析，使用十摺交叉驗證法，計算每一群的錯誤率並進行決策分析。

在進行分群後的決策樹分析之前，先對215筆的原始資料進行決策樹分

析觀察其分類正確率及分類情況，發現經過十摺交叉驗證之後，分類正確的不良品資料有170筆，而未能正確分類的不良品資料有45筆，總體分類正確率為79.06%，而本研究做分群的目的就是為了能讓相同性質的資料能夠一起分類，以達到決策樹分類正確率的提升，因此五群決策樹的正確率目標都比79.06%高。在未分群的情況下，系統判定最適合用來當作判別因子的為人員變項，其次為機器、方法及材料變項，而環境變項是最不適合當作判別因子的變項，與上一節分成五群的結果比較，發現人員和機器變項分別於主要及次要原因之間有了變動，以及環境和材料變項在次要和最不重要之間變動，初步證實了分群必要性，以下將以表格列出分類後的法則，提出分類正確率及探討每一群的不良情形，說明中僅列出出現分類錯誤情形之資料，以(N/X)表示，N代表此群中分類到這項不良的個數，X代表此不良個數中分群錯誤的筆數。

表4.30 第一群決策樹分類法則

1. IF 人員=碰觸污染膜面 AND 環境=配向膜預烤部風量不穩 THEN 判斷=縱不均
2. IF 人員=碰觸污染膜面 AND 環境=CELL 組立段環境異物 THEN 判斷=配向膜異物
3. IF 人員=碰觸污染膜面 AND 環境=ARRAY 製程環境異物 THEN 判斷=面板內異物
4. IF 人員=碰觸污染膜面 AND 環境=無影響 THEN 判斷=面板內異物
5. IF 人員=配向布裝配不平整 THEN 判斷=配向條紋 A
6. IF 人員=無影響 AND 材料=CF 來料不良 THEN 判斷=橫條紋不均
7. IF 人員=轉寫配向膜印刷用凸版未洗淨 THEN 判斷=雨狀不均
8. IF 人員=干涉刮傷配向膜 THEN 判斷=橫條紋不均
9. IF 人員=刮膠作業疏失 THEN 判斷=PKISS
10. IF 人員=靜電裝置未正確配戴使用 THEN 判斷=注入口附近黑色區塊
11. IF 人員=無影響 AND 材料=TFT 不良 THEN 判斷=SHOT
12. IF 人員=無影響 AND 材料=框膠材異常 AND 環境=CELL 組立段環境異物 THEN 判斷=半月 GAP
13. IF 人員=無影響 AND 材料=框膠材異常 AND 環境=無影響 THEN 判斷=面板四角發生 GAP

(資料來源:本研究整理)

表4.31 第二群決策樹分類法則

1. IF 方法=前洗壓入量不足 AND 材料=SPACER 受潮 THEN 判斷=配向膜異物
2. IF 方法=前洗壓入量不足 AND 材料=CF 原材異物 THEN 判斷=黑點
3. IF 方法=前洗壓入量不足 AND 材料= TFT 來料不良 THEN 判斷=黑點
4. IF 方法=配向膜吐出量設定問題 AND 人員=碰觸汙染膜面 THEN 判斷=縱不均
5. IF 方法=配向膜吐出量設定問題 AND 人員=轉寫刮刀裝配不平整 AND 材料=刮刀刀不平整 THEN 判斷=縱不均
6. IF 方法=配向膜吐出量設定問題 AND 人員=轉寫刮刀裝配不平整 AND 材料=CF 原材異物 THEN 判斷=橫條紋不均
7. IF 方法=配向膜吐出量設定問題 AND 人員=轉寫刮刀裝配不平整 AND 材料=SPACER 受潮 THEN 判斷=橫條紋不均
8. IF 方法=配向膜吐出量設定問題 AND 人員=轉寫刮刀裝配不平整 AND 材料=TFT 來料不良 THEN 判斷=橫條紋不均
9. IF 方法=轉寫速率設定 THEN 判斷=T 不均
10. IF 方法=配向參數設定 THEN 判斷=T 不均
11. IF 方法=無影響 AND 材料=刮刀刀不平整 THEN 判斷=膜面刮傷
12. IF 方法=無影響 AND 材料=CF 來料不良 THEN 判斷=T 不均
13. IF 方法=無影響 AND 材料=TFT 來料不良 THEN 判斷=T 不均

(資料來源:本研究整理)

從上一節的分析中得知，第一群主要不良類型為異物的產生及面板條紋不均等情形，與表4.30做對照，可看出能分類出來的不良現象也多是屬於這兩類。本群經十摺交叉驗證後，分類正確率達到82%，比未分群之前的正確率為高，顯示第一群之分類確實有達到分群的效果。而其中未完全正確分類的不良現象為縱不均(9.0/3.0)，橫條紋不均(2.0/1.0)及SHOT(9.0/5.0)等三類，可能的成因是這幾項不良現象中的關鍵變項與其他不良資料相似，例如在縱不均及橫條紋不均的人員變項成因都是碰觸汙染膜面及轉寫刮刀裝配不平整，而另一項也在第一群出現並有類似原因的T不均現象並沒能出現法則，但若使用其他的原因變項做分析則會導致分類正確率下降，導致系統判斷誤差更多，因此還是用原變項做分析。在第一群中除了這三種之外其他不良現象均能有效並正確的分類，因此建議品管人員若遇

到上述三種不良現象時，除了在此群之決策樹找規則外，尚需與其他群類似資料做比較，才能減少判斷錯誤之發生。

在上一節中得知，面板條紋不均、膜面刮傷及黑點的產生為本群主要的品質不良現象，與表4.32做比較，發現多數能分類出來的規則也都屬於這三類。本群經十摺交叉驗證後正確率達到82%，顯示第二群也確實達到分群使分類正確率上升的目的。其中未完全正確分類的為T不均(9.0/5.0)(9.0/2.0)(2.0/1.0)、黑點(2.0/1.0)、配向膜異物(3.0/1.0)、膜面刮傷(16.0/1.0)，可看出在本群中T不均這個不良現象出現多個判斷法則，並且其中有三條法則出現判斷誤差，建議日後若是品檢人員在此項不良現象出現時，應反覆多比對各條法則，並應考慮其他群之相關法則，才不致判斷錯誤，另外其他出現誤差的不良現象的筆數都只有一筆，因此若出現此項不良可按照此法則去判斷成因。

表4.32 第三群決策樹分類法則

1. IF 機器=機台配置不平整，吸著壓過大 THEN 判斷=PKISS
2. IF 機器=框膠爐溫離異 THEN 判斷=角殘
3. IF 機器=液晶皿汙染 THEN 判斷=注入口偏白
4. IF 機器=偏貼機構干涉發塵 THEN 判斷=偏光板不良
5. IF 機器=干涉刮傷配向膜 THEN 判斷=T不均
6. IF 機器=配向機震動 AND 環境=無影響 THEN 判斷=配向條紋 B
7. IF 機器=配向機震動 AND 環境=配向膜預烤部風量不穩 THEN 判斷=T不均

(資料來源:本研究整理)

由上一節第三群之資料得知，此群最多的不良情形為不均現象、條紋現象、角殘及PKISS，而這些項目在決策樹的運算中也都有分類出法則來。本群經十摺交叉驗證法後正確率約只有78%，比未分群前的79%略低，可知此群分群的效果並不怎麼顯著。細看各筆資料，發現有判斷誤差的法則為T不均(7.0/1.0)、配向條紋B(8.0/4.0)、角殘(5.0/2.0)等三項，其中T不均已在第二群中出現判斷錯誤情形，可知此項不良現象比較不能正確分類，品檢人員在分析時應更小心謹慎

而配向條紋B及角殘兩項不良情形在本群中都有過多的判斷錯誤情

形，可能原因為造成不良的變項與他筆不良資料過於相同，例如配向條紋B的機器變項不良原因為配向機震動，而這與縱不均和T不均的機器變項不良原因相同，導致系統使用機器變項做分類時判斷錯誤，由表中也可看出在此群中縱不均此不良現象並未被正確分類出一條法則，此時若使用材料變項做分類，在品質成因不同情況下，應可使決策樹分類較為正確，但若使用材料變項為判定因子，卻會使整體分類正確率下降，在此情形之下，本研究並不建議更改此群的判定因子，建議品檢人員在遭遇配向條紋B及角殘兩種不良現象時，應多加謹慎判斷，必要時採用全檢是較保險的檢測方式。

表4.33 第四群決策樹分類法則

1. IF 機器=轉寫/配向設備震動 THEN 判斷=配向條紋 B
2. IF 機器=機台配置不平整，吸著壓過大 THEN 判斷=面內 GAP
3. IF 機器=框膠爐溫離異 AND 材料=彩色濾光片原材異物 THEN 判斷=綠不均
4. IF 機器=框膠爐溫離異 AND 材料=框膠原材/異物 THEN 判斷=周邊 GAP
5. IF 機器=框膠爐溫離異 AND 材料=機板排版設計不易排氣 THEN 判斷=周邊 GAP
6. IF 機器=組立移載手臂偏移 AND 材料=TFT 或彩色濾光片原材寸法異常 THEN 判斷=漏光
7. IF 機器=組立移載手臂偏移 AND 材料=機板材質選用錯誤 THEN 判斷=漏光
8. IF 機器=組立移載手臂偏移 AND 材料=TFT 來料不良 THEN 判斷=斜視 SHOT
9. IF 機器=散佈機靜電異常 THEN 判斷=SPACER 打痕或凝集

(資料來源:本研究整理)

本群在上一節中得知筆數最多的不良資料為周邊GAP及綠不均，在表4.33中可看到兩種現象均能有效的得出決策法則，本群經十摺交叉驗證後的正確率為91.2%，比未分群前的正確率高，因此判定本群分群的效果良好。

在資料表現中，本群只有面內GAP(6.0/3.0)一條法則出現判斷誤差情形，詳細查看資料可發現，決策樹分析系統很有可能是把面內GAP及異物GAP兩種不良現象混為一種，可能原因為決策樹採用的是資料分割的概念(Breiman,1984)，透過對某連續數值屬性在資料集中出現的每一個屬性值逐一測試，以找出最佳臨界值，計算量十分龐大，而量度一個連續數值屬性的資訊貢獻值，以找出適當的臨界值的執行效率必定很低(童冠燁,1997)。

除此之外，當資料集含有超過兩個類別的情況下，計算資訊熵以選擇分割點時，必須考慮權重，使得分割點可能落入單一類別中，降低分割效果，造成所謂的不良分割點（bad point）（Fayyad and Irani, 1992）。欲解決此問題，必須加入更多判斷因子，以創造出更多決策樹上的節點，產生新的法則。但是此項誤差並不會影響判斷，因為兩種不良現象成因差不多，因此按照決策樹法則去檢測即可。

表4.34 第五群決策樹分類法則

1. IF 方法=ISO 洗淨功率過大 THEN 判斷=線不良
2. IF 方法=ARRAY 成膜參數設定 THEN 判斷=輝點
3. IF 方法=偏貼速度參數設定 THEN 判斷=靜電氣
4. IF 方法=剝膜機參數設定 THEN 判斷=彩色濾光片裂化
5. IF 方法=切裂機參數設定 AND 機器=散佈機靜電異常 THEN 判斷=SPACER 打痕或凝集
6. IF 方法=切裂機參數設定 AND 機器=機台上之夾管線角度異常或磨損 THEN 判斷=碰傷
7. IF 方法=點燈參數設定 THEN 判斷=殘像

(資料來源:本研究整理)

在上一節中得知第五群最多的不良現象為輝點、線不良及殘像，在表4.35的決策樹分類中都可以發現有法則出現，本群經十摺交叉驗證得到分類正確率達到100%，因此本研究判定此群的分群情況良好，而品檢人員在面對此群出現的不良現象時，也可以完全依照決策樹法則去進行檢查。

值得注意的是在本群中有一項不良現象配向條紋B並未能找出正確法則，可能原因為輸入向量的差異會造成SOM在映射時產生誤差，例如事先被歸類在同類別的兩個輸入向量，卻因為兩個輸入向量差距很大，因此在做分群時SOM就很有可能將兩者映射於不同的特徵空間(張斐章與張麗秋,2007)，此現象有學者提出了可以將SOM結合學習向量量化模式(Learning Vector Quantization ;LVQ)，將可有效的提升分群效果。(Suganthan,1999)(Baig et al.,2002)

4.3 本章小結

本研究成功的把蒐集到的215筆品管資料依照其特性分為五群，並得到可以用於判定不良成因的法則，總計為第一群共13條法則，第二群13條法則，第三群七條法則，第四群九條法則，第五群七條法則，共計四十九條法則，若日後品檢人員發現不良品質現象時，就可依據本研究所整理歸納出來的法則進行成因之探究。廠商在進行新廠擴建或是有新進人員時，可依據本研究之法則整理出檢查手冊進行不良原因之檢查或新進人員之訓練，更可讓整體運作快速進入狀況。但在任何產業中，經驗累積、技術進步、規模效應等都會產生學習效應(Pankaj,1985)(Jean-Paul,1985)，在TFT-LCD這種高競爭性的產業中亦然，本研究得出之法則有可能因為人員作業經驗之累積或是技術的進步而消失，也有可能產生新的法則，因此檢查手冊應視狀況更新，更新的方法即為重新蒐集不良現象的資料，運用本研究之流程重新得出新的法則，才能讓品質不斷的提升。

第五章 結論與建議

本研究為利用資料採礦方法針對台灣地區某家專業中小尺寸TFT-LCD製造廠商的品質問題做分析，使用資料採礦中的分群及決策樹對資料進行演算，期望能找出在出現品質異常時的分析法則，以利品質異常分析人員在日後發現品質不良現象時，只要依據本研究提出來的法則，就能快速找到問題的根源。本章將分為兩節說明，第一節對本研究做結論，分為方法上的結論及資料處理後的結論兩部分，第二節將討論後續的研究發展。

5.1 結論

5.1.1 方法面結論

- 1.二階段分群法方面:本研究為了得出正確的分群數目，採用了二階段分群法，也就是先使用階層式分群法中的華德法決定分群數目，接著使用K平均數來做分群後的驗證。研究結果顯示，本研究所採用的資料適合分為五群，再於K平均數方法中決定各群應注重的判定因子，顯示出環境因子最不適合拿來當作判定因子，其可能因為環境變項的不良種類最少導致較難分類，也有可能因為K平均數方法對資料分群不均所致。
- 2.自我組織映射圖網路分群方面:除了前述文獻提到SOM的分群效果在資料不確定情形下較正確外，Mangiameli等人(1996)也曾提到，由於資料中離群值對於向量學習過程的影響有限，自我組織特徵映射圖上的拓撲網路點對於原始資料的隨機變異之敏感度較低，因此自我組織特徵映射圖處理不完美資料的效能較優越，而Vesant & Alhoniemi (2000)亦證實了其分群精確度也能提升；因此本研究再以SOM方法針對資料進行分群，而為了使判斷具有公平性，本研究一樣將資料分為五群，結果顯示SOM的分群效果確實比K-means分群來的好，可見SOM方法確實能減低資訊中的雜訊，將資料的真實樣貌表現出來，而另外一方面，經過分群後判定因子變項的重要性也隨之改變，最重要的變項由人員變項轉變為機器變項，因此分群確實較能正確的判讀資料。
- 3.決策樹分析方面:本研究以SOM所得到的分群資料進行決策樹分析，試圖找出不良產品發生的法則，結果顯示在分群後進行決策樹分析除了在第三群的正確率不及未分群前的正確率外，其餘都能有效的提升決策樹分類的正確性，而第三群未能達到目標值可能是因為目標資料有雜訊導

致，但也沒有偏差太多，因此本研究之決策樹分析資料也可供廠商做品檢上的參考。

5.1.2 分析面結論

- 1.在結果中顯示，最重要的決策變項為機器變項，相對的也就是說在機器方面產生的不良現象最多。從原始資料中得知，在機器變項中最常出現的就是配向膜相關問題，例如設備干涉刮傷配向膜、配向膜機器震動等，在LCD面板中的配向膜具備讓液晶分子依設定方向排列的功能，塗佈於TFT基板及CF基板兩側，直接與液晶分子接觸，可以說為LCD之關鍵材料，而其中的聚醯亞胺（Polyimide；PI）更是決定面版品質的重要材料，但是台灣的面板廠目前仍多仰賴日本進口此種原料，關鍵材料無法掌握實為一大隱憂。目前國內已有廠商在研發相關技術，如奇美、達興等，建議個案廠商應對於此關鍵材料的來料狀況多加掌握，與握有關鍵技術的廠商發展合作關係，應可大幅提高良率。
- 2.在本研究中另一個較重要的變項為方法變項，在此變項中大多為一些機器參數的設定，如加壓封止參數設定、切裂機參數設定等，而大多是溫度、壓力方面的問題，在這部分就需要針對發現的問題做反覆的測試，重新設定最適合的參數來提高品質。
- 3.本研究提出的決策樹法則結論，適合用來做為日後品質異常預警之用，因為從資料中可看出，法則中得到最重要及最有可能產生最後不良現象的原因，若發生類似現象就有可能會出現最後不良結果，留意不良成因的出現可以預防不良現象的產生。

5.2 未來研究與建議

本研究在實證分析的部分，提出了一個可以快速檢驗TFT-LCD中面板組裝製程品質問題的模式，但是在實際分析上還有許多問題需要探討，也有許多方式可以做進一步的分析，現分述如下：

- 1.資料筆數及蒐集時間:在本研究中可以發現有些不良現象未能歸納出判斷法則，可能因為本研究所探討的個案公司其運作時間不長，並未有大量資料，雖已達可以分析的標準，但建議未來可再蒐集更多資料筆數做分析，或許可產生更多法則。
- 2.大尺寸面板的應用:大尺寸與中小尺寸面板檢測方式有些許不同，中小尺寸目前是以人工目視檢測為主，而大尺寸面板廠商多已使用光學檢測的

技術檢驗，雖已比人工目視的正確率高出很多，但是若能更增進檢驗速度，將可更精簡成本。且大尺寸面板可說是目前台灣TFT-LCD的主力產品，提升良率將可提升整體競爭力，建議可以大尺寸面板之資料來做分析。

- 3.更細部的品質分析:本研究所採用的資料為面板組裝製程的整體品質不良現象，可以巨觀的看出品質問題，但是在整體製程中還有許多更細部的過程，例如最關鍵的配向膜製程中，在PI液轉寫到基板上之後，需要利用熱板(hot plant)進行預烤，接著再用燒成爐進行熱處理，這都還在本研究資料蒐集範圍，但是在燒成方式中又有遠紅外線加熱(infrared heater)及熱風循環加熱(heated air circulating)兩種製造方式(田民波,2008)，或許又可產生更多判斷法則，因此建議可以針對更細部的製程做分析。
- 4.不同分群與分類方法:在本研究中可發現有些不良現象並未能出現法則，而有些則出現多個法則，在判斷上可能會出現誤差，未解決這些問題，可運用其他資料採礦方法，在分群方面有統計方法中的模糊C平均數(Fuzzy C-means)，類神經網路中的適應共振理論(Adaptive Resonance Theory)等，都有各自不同的特性，未來研究可考慮使用這些方法來做分析;在分類方法方面，Olaru and Wehenkel (2003)提出柔性決策樹(soft decision trees)，將模糊觀念建立在決策樹的分支上，模糊邏輯由於模糊集合的彈性，可以改善一般決策樹在分類上較高的變異，另外Cordon et al.(2001) 利用模糊邏輯 (fuzzy logic)的近似搜尋法，結合決策樹之範例學習，可以解決並能有效地提取出模糊規則，選取合適的節點，提供有用的資訊，利用上述方法應可有效改善本研究分類錯誤的情形。

參考文獻

一、中文部分

王人星，2004。製造貨批量變動之製造系統模擬與資料挖礦以一面板廠為實證研究，國立清華大學工業工程與工程管理學系碩士論文。

王昭琪，2002。TFT-LCD產業之競合與經營型態分析，工研院產業經濟與資訊服務中心。

王綺年，2008。國家角色與產業發展之連結：比較台灣與韓國之薄膜液晶顯示器產業，東亞研究，39卷1期

田民波，2008。TFT LCD面板設計與構裝技術，五南圖書出版公司。

朱正中，2004。台灣TFT-LCD產業的垂直整合策略分析—自利潤池的角度切入，國立交通大學管理科學研究所碩士論文。

李秀玉，1998。應用賽局理論分析我國薄膜電晶體液晶顯示器產業之競爭策略，國立交通大學科技管理研究所碩士論文。

李俊昇，2003。液晶面板組裝廠批量製程派工法則之設計，交通大學工業工程研究所碩士論文。

林鼎浩，2003。建構半導體製程資料挖礦架構及其實證研究，碩士論文，清華大學工業工程與管理研究所。

李培瑞，2002。半導體製程資料挖礦架構、決策樹分類法則及實證研究，碩士論文，清華大學工業工程與管理研究所。

呂俞德，2006。TFT-LCD缺陷檢驗之搜尋與分類模式，國立台北科技大學工業工程與管理系碩士論文。

吳明隆、涂金堂，2005。SPSS與統計應用分析，五南圖書出版公司。

易任、葉惠中，1990。主成分分析與群集分析應用於雨量空間分布之研究，易任論文集，549-563。

林立偉，2006。台灣中小尺寸顯示面板公司之創新矩陣策略分析，國立交通大學管理學院碩士在職專班科技管理組碩士論文。

林震岩，2007。多變量分析 SPSS的操作與應用，智勝文化事業有限公司。

洪煌龍，2004。資料探勘在社區大樓電信服務市場之應用，國立成功大學工業與資訊管理學系專班碩士論文。

洪瑜君，2006。以資料探勘技術識別造成彩色濾光片製程缺陷之群聚現象，國立成功大學工業與資訊管理學系專班碩士論文。

柯長泰，2005。薄膜電晶體液晶面板生產之製令規劃，國立雲林科技大學工業工程學系碩士論文。

高士傑，2001。區域性設計兩型建立方法之研究，國立台灣大學土木工程研究所碩士論文。

莊智偉，2007，液晶面板廠機台異常分析應用資料採礦之研究，輔仁大學應用統計學研究所碩士論文

陳志忠，2001。液晶顯示器的像素點缺陷與亮度均一性之自動化檢測，私立中原大學機械工程研究碩士論文。

陳響亮、張頌榮，2005。TFT-LCD面板之點線瑕疵自動化檢測系統，第二十二屆中國機械工程學會全國學術研討會論文集。

陳坤盛、王靖欣、陳細鈿，2004。應用標準差之MAIC流程改善TFT-LCD面板之製程品質，智慧科技與應用統計學報,第二卷,第二期。

郭信宏，2007。應用資料探勘技術於面板檢測實證研究，國立中央大學工業管理研究所碩士在職專班碩士論文。

黃俊英，2003。多變量分析，華泰書局。

黃瓊芬，2003。應用關聯性規則探勘於股市時間序列分析—以TFT-LCD產業股價行為為例，國立交通大學資訊管理研究所碩士論文。

張紹勳、張紹評、林秀娟，1993。SPSS For Windows多變量統計分析，松崗電腦圖書資料股份有限公司

張華城、張育銘，2000。以SOM網路建構造形認知模擬系統可行性研究，2000年技術與教學研討會，明志技術學院

張嘉麟，2000。台灣薄膜電晶體液晶顯示器產業核心能力與競爭優勢分析，國立交通大學科技管理研究所碩士論文。

曾彥馨，2003。應用機器視覺於TFT面板之表面瑕疵檢測與分類，私立元智大學工業工程與管理研究所碩士論文。

葉怡成，2003。類神經網路模式應用與實作，儒林圖書有限公司。

鄒民欽，2003。TFT-LCD製程探討-以模組廠製程的改善為案例，國立清華大學工業工程學系碩士論文。

童冠燁，1997。以歸納學習法探討彈性製造系統動態排程之研究，國立成功大學工業管理研究所博士論文。

鄭春生、周雅君，2007。以資料探勘為基建構偏光板品質異常診斷系統，中華民國品質學會第43屆年會暨第13屆全國品質管理研討會。

蔣漢旗，2002。台、韓TFT-LCD製造發展策略比較分析之研究，國立交通大學經營管理研究所碩士論文。

鄭惠珊、鄭景俗，2007。資料探勘技術評估-製造業供應商績效ICIM第十八屆國際資訊管理學術研討會。

錢志豪，2003。建構液晶顯示器(LCD)色彩偏差瑕疵之自動化視覺檢測系統之探討，私立朝陽科技大學工業管理研究所碩士論文。

簡禎富、李培瑞、彭誠湧，2003。半導體製程資料特徵萃取與資料挖礦之研究，資訊管理學報，第十卷，第一期。

謝仲為，2002。先進規劃與排程系統應用於 TFT-LCD 產業之研究，東海大學工業工程研究所碩士論文。

謝邦昌，2005。資料探勘與商業智慧 - SQL Server 2005，鼎茂圖書出版股份有限公司。

謝昌倫，2003。以資料採礦方法辨認半導體晶圓圖的錯誤樣式，私立淡江大學統計系應用統計研究所碩士論文。

魏永康，2003。臺灣TFT-LCD產業之競爭策略分析，國立台北大學企業管理學系碩士論文。

羅華強，2005。類神經網路—MATLAB的應用，高立圖書有限公司。

二、英文部分

Abidi, S. S. R. and Ong, J., 2000. "A Data Mining Strategy for Inductive Data Clustering: A Synergy Between Self-Organizing Neural Networks and K-Means Clustering Techniques," *TENCON Proceedings*, Vol. 2, pp.568-573.

Andrew, K. and Christian K., 2001, "Data Mining of Printed-Circuit Board Defects", *IEEE Transactions on Robotics and Automation*, Vol. 17, No. 2, April, pp. 191-196.

Andrew, K., 2000, "Decomposition in Data Mining: An Industrial Case Study", *IEEE Transactions on Electronics Packaging Manufacturing*, Vol. 23, No. 4, October, pp. 345-353.

Baig, M.H.; Rasool, A.; Bhatti, M.I.;2002. "Classification of electrocardiogram using SOM, LVQ and beat detection methods in localization of cardiac arrhythmias. Proceedings of the 23rd Annual EMBS International Conference. October 25-28.

Balakrishnam, P. V., Cooper, M. C., Jacob, V. S., and Lewis, P. A., 1994. "A Study of the Classification Capabilities of Neural Networks Using Unsupervised Learning - A Comparison K-means Clustering," *Psychometrika*, Vol. 59, No. 4, pp. 509-525.

Berry, M. and Linoff, G., 2004, *Data Mining Techniques for Marketing, Sales and Customer Support*, John Wiley, New York.

Breiman, L., Friedman, J., Olshen, R., & Stone, C., 1984. *Classification and Regression Trees*, Wadsworth International Group, Belmont, California.

Bradley, P. S. and Fayyad, U. M., 1998. "Refining Initial Points for K-means Clustering," *Proceedings of the Fifteenth International Conference on Machine Learning*, San Francisco, pp. 91-99.

Cabena, P., Handjinaian, P.O., Stadler, DR.J., Verhees, J. and Zanasi, A., 1998. "Discovering Data Mining from Concept to Implementation," Prentice Hall.

Chapman P., Clinton J., Khabaza T., Reinartz T., Wirth R., 1999. *The CRISP-DM Process Model*

Chi, Z., Peter, C. N., Weimin, X., Thomas, M. T., 2001, "An Intelligent Data Mining System for Drop Test Analysis of Electronic Products", *IEEE Transactions on Electronics Packaging Manufacturing*, VOL. 24, NO. 3, JULY, pp. 222-231.

- Cordon, O., Herrera, F. and Villar, P., 2001, Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base, *IEEE Transactions on Fuzzy System*, 9(4), 667-674.
- Dan, B. and Armin, S., 2002, "Data Mining for Improving a Cleaning Process in the Semiconductor Industry", *IEEE Transactions On Semiconductor Manufacturing*, Vol. 15, No. 1, February, pp. 91-101.
- Elmasri and Navathe., 2000., *Fundamentals of Database Systems*, PEARSON EDUCATION, Inc.
- Ester, M., kriegel, H. P., Sander, J., and Xu, X. ,1996. "Density-Based A1gorithm for Discovering Clusters in Large Spatial Databases With Noise," *In Proceedings Conference on Knowledge Discovery and Data Mining*, Portland, pp. 226-231.
- Fayyad, U. M., & Irani, K. B. ,1992. On the handling of continuous- valued attributes in decision tree generation. *Machine Learning*, Vol.8, No.2, pp.87-102, Kluwer Academic Publishers, Massachusetts.
- Forgy, E. ,1965. "Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications," *Biometrics*, Vol. 21, p. 768.
- Garcia-Fornes,A.,Hassan H. and A. Crespo,1996. "Strategies for Scheduling Optional Tasks in Intelligent Real-Time Environments," *Journal of Systems Architecture*, 42, 391-407
- Hadjinian, P., Stadler, P., Verhees,R., Zanasi, J.,1997. *Data Mining from Concept to Implementation*. , Prentice-Hall: Englewood Cliffs, NJ.
- Han, J. and Kamber. M.,2004.*Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.
- Iivarinen, J.,1994."Visualizing the Clusters on the Self-Organizing Map," in Proc. Conf. Artificial Intel Res.Finland, C. Carlsson, T. Järvi, and T. Reponen, Eds. Helsinki, Finland, pp 122–126.
- Jain, A.K.,2001, "Data Clustering: a review", *ACM Computing Surveys*, Vol.31, Issue3, pp.264-323.
- Jeong, B, Kim, SW and Lee, YJ , 2001. An assembly scheduler for TFT LCD manufacturing. *Computers and Industrial*, 41(1):37-58
- Jean, P.S.,1985 "The Uses and Abuses of Experience Curves," *Long Rang Planning*, Vol.18, No.1, pp.65,
- Kantardzic .,2003. *Data mining: Concepts, models, methods, and algorithms*. Piscataway, NJ: Wiley-IEEE.
- Kennedy R. L and Lee Y. and Van Roy B. and Reed C.D. and Lippman R.P.,1998. *Solving Data Mining Problems Through Patern Recognition*, Upper Saddle River: NJ:Prentice Hall
- Kohonen T., Kaski, S., Lagus , K., Salojävi, J., Honkela, J., Paatero, V., & Saarela,A ,2000. Self-organization of a massive document collection. *IEEE Trans. Neural Networks*, 11(3): 574-585.
- Kohonen, T., 1995. *Self-organizing Maps*. Berlin: Springer.

- Kraaijveld, M.A., 1995, "A Nonlinear Projection Method Based on Kohonen's Topology Preserving Maps," *IEEE Trans. Neural Networks*, vol. 6, pp 548–559.
- Mangiameli, P., Chen, S. K., and West, D., 1996. A Comparison of SOM Neural Network and Hierarchical Clustering Methods, *European Journal of Operation Research.*, Vol.93, No. 2, Olaru, C. and Wehenkel, L., 2003, A complete fuzzy decision tree technique, *Fuzzy Sets and Systems*, 138, 221-254.
- Pankaj, G., 1985. "Building Strategy on the Experience Curve," *Harvard Business Review*, p.145~146
- Pena, J. M., Lozano, J. A., and Larranaga, P., 1999. "An Empirical Comparisons of Four Initialization Methods for the K-means Algorithm," *Pattern Recognition Letters*, Vol. 20, pp. 1027-1040.
- Pankaj, G., 1985 "Building Strategy on the Experience Curve," *Harvard Business Review*, pp.145~146
- Richard J. Roger, Michael W. Geatz., 2003. *Data Mining: a tutorial-based primer*, First edition, Boston, Addison Wesley, pp. 148-164.
- Robert Hecht-Nielsen., 1990. "Neurocomputing," Addison-Wesley, Reading, MA,
- Rodammer, F.A. and K.P. White, 1988. "A Recent Survey of Production Scheduling," *IEEE Transactions on Systems, Man, and Cybernetics*, 18(6), 841-851
- Safavian, S. R., & Landgrebe, D. (1991). A Survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man and Cybernetics*, Vol.21, No.3, pp.660-674.
- Sharma, S., 1996. "Applied Multivariate Techniques," Wiley, pp212-216
- Jiang, D., Tang, C. and Zhang, A., 2004. "Cluster Analysis for Gene Expression Data: A Survey," *IEEE Trans. on Knowledge and Data Engineering*, vol.16, no. 11, pp.1370-1386.
- Shin, HJ and Leon, VJ, 2004. Scheduling with product family set-up time: an application in TFT LCD manufacturing. *International Journal of Production Research*, 42(20) : 4235-4248
- Suganthan, P. N., 1999. Hierarchical Overlapped SOM's for Pattern Classification. *IEEE Transactions on Neural Networks*, 10(1), 193-196.
- Sun, Y., Zhu, Q., and Chen, Z., 2002. "An Iterative Initial-Points Refinement Algorithm for Categorical Data Clustering," *Pattern Recognition Letters*, Vol. 23, pp. 875-884.
- Toba H., 2004. A Tight Flow Control for Job-Shop Fabrication Lines With Finite Buffer. *IEEE Transactions on Automation Science and Engineering* : Accepted for future publication, pp(99):1-6
- Tseng, Y.H. and Tsai, D.M., 2006, Using Independent Component Analysis Based Process Monitoring in TFT-LCD Manufacturing, *Journal of the Chinese Institute of Industrial Engineers*, Vol. 23, No. 3, pp. 262-267
- Ullsh, A. & Siemon, H.P., 1990. "Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis," in *Proc. Int. Neural Network Conf. Dordrecht*, The Netherlands, pp 305–308.
- Vesanto, J., Alhoniemi, E., 2000. Clustering of the Self-Organizing Map, *IEEE Transactions on Neural Networks*, Vol.11, No.3, pp.586-600.

Vetter, C., 1993, "Self-Organizing-Feature-Maps versus Statistical Clustering Methods: A Benchmark", *Research Report 0994*, University of Marburg, FG Neuroinformatik & Künstliche Intelligenz.