

臺北醫學大學醫學資訊研究所

碩士論文

Taipei Medical University

Graduate Institute of Medical Informatics

Master Thesis

指導教授：邱泓文 Chiu, Hung-Wen

蔣以仁 Chiang, I-Jen

以叢集分析技術探討病患就診屬性與看診時間之關係
Clustering technology for predicting physician service time
by using outpatient's attributes and hospital resource utilization

研究生：黃仁貴(Huang, Jen-Kuei) 撰

中華民國九十一年六月

June , 2002

誌 謝

還記得 2000 年人類基因體計畫成果剛公佈時，學術界似乎有一股狂熱，只要能沾上邊的，都想湊上一腳。本所李友專所長，亦極力的投入此領域，強調跨領域是未來的趨勢，同時邀請榮陽團隊的多位教授，前來本所演講基因相關課題，及成立生物資訊研究中心，並架設完成叢集電腦架構，在此感謝李所長之用心，與對本所學生不時之督導與深切之期望。在當時，對電腦熟悉的我加上又在醫院工作，找個與疾病相關的基因題目，應當是最恰當不過了。然而研一結束後，每位同學都在趕著論文計畫書時，才發覺自己在基因之領域上，一直無法跨入並提出適當的研究題材。所幸在資料探勘之課程研討中，我很感激蔣老師之指導，還有一群共同研究的伙伴們，在相互討論中給予我的建議與指導，讓我找到一個適當的題目。

在此要感謝我的指導老師—邱泓文教授，在他的指引之下，讓我由茫茫的研究領域裡找出一個目標，並不厭其煩的指導我，指證我的缺失，如今才有辦法於百忙中完成此論文。另外，在資料探勘方面，我要謝謝我的另一位指導老師—蔣以仁教授，由於他親切又認真的教導，還提供 On call 之線上服務，使我在理論與應用上能對資料探勘有更深入的瞭解。最後感謝劉德明教授，分別對於論文計畫書審查以及論文口試時的寶貴意見，使本論文品質能更上一層。

另外，我也要感謝我的太太對我唸書的支持、鼓舞與叮嚀，並讓我不必為家中雜事擔心，能夠心無旁騖，好好完成研究所學業，最後感謝醫資所的所有老師這兩年來的培育，使我成長了更多。

黃仁貴

於臺北醫學大學醫學資訊研究所

2002/06/16

目 錄

	頁數
標題.....	i
審定書.....	ii
上網授權書.....	iii
國科會授權書.....	iv
誌謝.....	vi
目錄.....	vii
表目錄.....	x
圖目錄.....	xi
中文摘要.....	xii
英文摘要.....	xiv
第一章 緒 論.....	1
1.1 研究背景與研究動機.....	1
1.2 研究目的.....	2
1.3 研究範圍與假設.....	2

第二章	文獻探討.....	4
2.1	醫療行銷.....	4
2.2	醫療品質指標與病患滿意度.....	5
2.3	醫院預約方式與制度.....	7
2.3.1	醫院預約系統.....	8
2.3.2	醫院實施預約制度之優缺點.....	9
2.3.3	預約制度在各層級醫院實施方式與結果.....	9
2.3.4	門診等候時間各屬性之資料分析.....	11
2.4	資料探勘.....	12
2.4.1	Data Mining 的功能.....	13
2.4.2	資料探勘(Data Mining)建置步驟.....	15
2.4.3	資料探勘與資料庫倉儲.....	16
2.5	等候心理學.....	18
2.6	病患等候時間與相關改善研究.....	20
第三章	叢集.....	22
3.1	叢集之基本概念.....	22
3.2	叢集建置之階段工作.....	23
3.2.1	模式表現 (Pattern representation).....	23
3.2.2	相近模式(Pattern proximity).....	24
3.2.3	叢集分類(Clustering taxonomy).....	25
3.2.4	評估叢集的結果(Assessment of output).....	26
3.3	叢集技術的分類 (A Taxonomy of Clustering Approaches).....	26
3.4	叢集的演算法.....	27
3.4.1	階層式叢集演算法(Hierarchical Clustering Algorithms).....	28
3.4.2	分割式叢集演算法(Partitional Algorithms).....	29

第四章	研究設備與方法.....	36
4.1	資料收集.....	36
4.2	研究材料與設備.....	37
4.3	研究方法與架構.....	37
4.4	研究步驟與流程.....	39
4.5	Index Miner 之使用.....	40
第五章	研究結果.....	42
5.1	門診病患資料描述與樣本資料分析.....	42
5.2	門診病患就診屬性與醫師看診時間之關係.....	44
5.3	以資料探勘之 EM 法分析病患屬性與看診時間之關係.....	51
5.4	叢集法之參數設定.....	51
5.5	叢集分析之結果.....	54
第六章	結論與討論.....	60
6.1	研究結論與討論.....	60
6.2	未來研究方向.....	61
	參考資料.....	64
	中文文獻.....	64
	英文文獻.....	66

附 錄 68

附錄一 D63 醫師以 Index Mining 軟體所得到 EM 之結果 68

表目錄

表 2-1 醫療服務品質面之 KPI.....	6
表 2-2 門診病患認為台北榮總應該改進的地方.....	7
表 2-4 門診看診流程之等候時間分析表.....	12
表 5-1 病患看診時間與看診時段的關係.....	45
表 5-2 門診病患人數與星期別之關係.....	47
表 5-3 病患各屬性與醫師看診時間之關係.....	50
表 5-4 個別醫師新病患與初、複診病患資料.....	52
表 5-5 看診時段與人數資料.....	52
表 5-6 一般內科 A86 醫師之 EM 叢集結果.....	55
表 5-7 神經內科 A78 醫師之 EM 叢集結果.....	56
表 5-8 小兒科 C28 醫師之 EM 叢集結果.....	57
表 5-9 一般外科 D63 醫師之 EM 叢集結果(其原始之叢集資料，請參考附件一).....	58
表 5-10 婦產科 B89 醫師之 EM 叢集結果.....	59

圖目錄

圖 2-1 資料探勘建置步驟	15
圖 2-2 一個典型資料倉儲的架構	17
圖 3-1 對於身高與體重與性別叢集的結果	23
圖 3-2 叢集技術的分類	27
圖 3-3 單一連結法凝聚步驟之圖例	28
圖 3-4 為 A 和 B 兩常態分配圖及其分佈比率	32
圖 4-1 性別與醫師看診時間之關係圖	37
圖 4.2 以叢集方式評估醫師看診時間之架構	38
圖 4.3 以叢集方式評估醫師看診時間研究流程	39
圖 4-4 INDEX MINER 資料載入方式	40
圖 4-5 參數設定	41
圖 4-6 EM 之參數設定與執行	41
圖 5-1 為各科看診人數直條圖	42
圖 5-2 為各醫師每月看診人數之直條圖	43

論 文 摘 要

論文名稱：以叢集分析技術探討病患就診屬性與看診時間之關係

臺北醫學大學醫學資訊研究所

研究生姓名：黃仁貴

畢業時間：九十學年度第二學期

指導教授：邱泓文 臺北醫學大學醫學資訊研究所（助理教授）

協同指導教授：蔣以仁 臺北醫學大學醫學資訊研究所（助理教授）

目前各醫院均秉持提供高服務品質的理念，而在各項服務品質的指標中，對病患最直接的感受就是等候看診時間，故預約看診時間與實際就診之等待時間長短，將直接反應於醫院之管理績效與對病患之服務品質。如果能依病患就診屬性與醫院所提供之服務，建立一預測醫師看診時間之醫療行為模式，作為看診等候時間之依據，以縮短病患就診等候時間，除了能提昇醫院經營效率強化其個人化服務品質外，此亦成為民眾選擇就醫醫院的重要考量因素。在醫院之立場，如何妥善規劃有限之醫療資源，做好醫病關係，以建立社區醫學之願景，亦是各層級醫院之重大課題。

然而就診病患之多樣性與不確定性常常造成 醫院排程之失調，使預約制度之看診時間可靠性大打折扣，此現象不止令醫院方面產生極大的困擾，亦令已生病的病患在身心上蒙受一層痛苦的就醫經驗。由於資訊的進步，電子病歷之實施雖可有效改善病歷調閱時間上問題，然而並無法有效改善病患就診等候看診時間上之問題。雖然有些醫院對同一醫師以等距之看診時間，做為病患看診時間之基礎，而面對此龐大之資料與眾多之變數，其等候時間依然無法有效改善。

本研究依病患就診與就醫之屬性，首先透過資料庫取得門診病患之資料，包括：看診日期，新病患、初、複診病患，看診時段（早、中、晚），醫師，年齡，性別與疾病碼

等。此資料經過濾與清除後，以資料探勘之叢集(Cluster)技術，探討個別醫師看診時間之叢集模式，此模式依屬性之相似特性加以分類，區分出不同之病患族群與其所佔人數之比率，以及每一族群下醫師看診所花費之時間。依此時間以做為設定病患看診時間依據，此將有助於建立一個有效且正確之病患看診時間預測模式。

Abstract

Title of Thesis : Clustering technology for predicting the physician service time by using outpatient's attributes and hospital resource utilization

Author : Huang Jen-Kuei

Thesis advised by : Chiu, Hung-Wen ; Chiang, I-Jen

Taipei Medical University,

Graduate Institute of Medical Informatics

Waiting time is a major performance indicator in measuring outpatients' satisfaction of health care services. However, factors which influence waiting time, especially for outpatients waiting for physician service, are more complicated and uncertain.

Outpatient clinic scheduling was studied extensively between 1950 and 1980. The objective of this study is to build a model referencing patient attributes and hospital resource utilization to predict medical service time for outpatients. Outpatient attributes includes first or subsequent visit, gender, age and disease. Hospital resource utilization includes clinic, shift, week and doctors. Based upon the above attributes we will use data mining tools to build a model for predicting physician service time.

This model - grouping by cluster technology - can be used in the outpatient reservation system to decrease outpatient waiting time and improve the patient-physician relationship. It also can help administrators to allocate equipment and human resources to improve total service quality.

第一章 緒 論

1.1 研究背景與研究動機

目前各醫院均秉持提供高服務品質的理念，而在各項服務品質的指標中，對病患最直接的感受就是等候看診的時間，故預約看診時間與實際就診之等待時間長短，將直接反應於醫院之管理績效與對病患之服務品質，而對服務病患而言，病患「感受」與「實際」的等候時間有顯著差異，且等候時間越長差距越大。病患在候診時期細數著看診號，此段時間造成心理錯覺，雖然是短短的時間卻感覺相當的漫長，感受的時間影響病人滿意度比實際的時間強(劉敏玲，2001)。如果能依就診的屬性不同，剛好即時提供病患所需的服務，除了能提昇醫院經營效率強化其個人化服務品質外，此亦成為民眾選擇就醫醫院的重要考量因素。在醫院之立場，如何妥善規劃有限之醫療資源，做好醫病關係，以建立社區醫學之願景，亦是各層級醫院之重大課題。而充份瞭解就醫病患之屬性，對醫院資源規劃與整體運作績效是息息相關的。諸如：在人力方面有充裕時間安排與調度各時段醫護與行政人員。在空間方面，可以疏通病人淨化就診品質。在醫療品質方面，可依病患之特性，增設特別門診及專科醫師。

然而就診病患之多樣性與不確定性，常常造成醫院排程之失調，使預約制度之看診時間可靠性大打折扣，此現象不只令醫院方面產生極大的困擾，亦令已生病的病患在身心上均蒙上一層痛苦的就醫經驗。而在國人依然迷戀於大醫院就診之情形下，亦造成各層級醫院之平均等候時間依醫院大小而不同，如醫學中心平均等待看診時間為 99 ± 31 分，區域醫院為 72 ± 25 分，地區醫院為 36 ± 10 分，診所為 26 ± 16 分(吳佩璟，2000)。雖然電子病歷之實施可有效改善病歷調閱之時間問題，然並無法有效改善病患就診等候就診時間。其中影響之因素包括醫院之看診科別、醫生、看診時段及病患之初、複診與主診斷等因素。而面對此龐大之資料與眾多之變數，並無法用目前各醫院所使用之等距法以做為看診時間之基礎。

本研究依病患就診與就醫之屬性，包括：看診日期，新病患、初、複診病患，

看診時段別(早、中、晚)，醫師別，年齡，性別，疾病碼等，以資料探勘之叢集(Cluster)技術，直接透過資料庫取得門診就診資料，尋求醫師看診之叢集模式，以做為設定看診時間依據，進而改善醫病關係提高醫療服務品質。

1.2 研究目的

根據研究動機之描述，因病患之多樣性與疾病之不確定性，使一般醫院無法有效掌握醫師看診時間，導致病患預約後，等候醫師看診時間漫長，因而使醫院之醫療品質無法有效提升。本論文之目的在於以叢集的理论基礎，依病患就診之屬性與就醫之屬性，建立一預測醫師看診時間之醫療行為模式，並藉由此模式所提供之知識，以達到下列功能：

- 依據病患就診屬性之叢集結果，提供醫師看診時間預測模式，以作為病患預約就診時，依病患就診屬性，提供正確看診模式之時間，以降低病患等候看診時間，並提高醫療品質。
- 提供醫院對每一位醫師合理門診量之設定參考。

1.3 研究範圍與假設

• 研究範圍

本研究將以某區域教學醫院，一個月之門診看診病患資料為主，為利於研究與說明，本研究將有關就診屬性：如病患本身、醫院提供病患看診之相關資源與醫師等，之細部項目與定義分述如下：

- 病患種類：用以區分病患就診特性包括：新病患、初診及複診病患三種。
- 新病患：病患首次來院就診，該院尚無該病患之基本病歷資料。
- 初診：指該病患曾在該醫院就診，有就診病歷紀錄與基本病歷資料之病患，但此次就診並非在與上次相同之科別與醫師下看診。

- 複診：指該病患曾在同一醫院同一科別之同一醫師就診，有就診病歷紀錄之病患。
 - 時段：醫院提供病患每天看診之起訖時間，如早、中與晚上三個時段。
 - 星期：醫院提供病患看診日期，如週一至週六或週日等。
 - 看診醫生：病患掛號後替病患看病之醫師。
 - 看診科別：病患掛號等待看診之科別。
 - 看診人數：病患掛號後等待該醫師看診之總人數。
 - 醫師看診時間：指病患進入診間後，醫師對該病患開始看診至結束之時間。
 - 主診斷：病患就診之主要疾病別，以 ICD-9 碼為主。
- 研究假設

本研究假設每一位病患依不同之就診屬性，包括病患屬性與就醫屬性之不同，看診時間將會有所不同，但卻有不同程度之相似性存在。

第二章 文獻探討

2.1 醫療行銷

行銷是確認顧客需求、決定目標市場、設計產品與服務以滿足顧客的一種企業機能。醫療行銷則包括了以實際醫療行為與心理上之形象感覺為定位準則。醫院的形象定位不僅是醫療行銷的重要課題，更是一種策略管理，有助於達成組織的多重目標與提昇醫院的競爭優勢。為瞭解醫院形象之重要性，在張櫻淳(1996)研究論文中，採以問卷調查之方式，針對某台北六家醫院之家醫科就診病患為研究對象，其結果發現五個醫院形象構面之重要性排名依序為 1.醫療品質 2.醫院的整體觀感 3.醫院及醫師之名聲 4.等候時間的長短 5.就醫便利性，可知等候時間為構成醫院形象之重點，且研究中指出消費者對醫院形象的認知與其就醫選擇具有相關性，因此等候時間較短的醫院有助於民眾在就醫選擇時成為優先考量。而研究中並指出，等候時間對於醫療品質則是呈市場空隙，因此可據此發展行銷定位策略，所以看診流程時間亦可成為醫療差異化行銷之重點。

同樣在其他評估醫院形象的構面研究中，也發現等候時間是構成醫院形象及民眾和醫院管理者評估的要項之一。根據蔡雅芳(1996)之研究，以台中某四家教學醫院之門診病患及行政管理為研究對象，採問卷調查之研究方法，在探索醫院形象及病患就醫態度之構面結果發現，評估醫院形象的構面為：1.就醫可近性 2.醫療品質 3.院內氣氛 4.等候時間，可知等候時間對於民眾與管理者，為在評估醫院形象和就醫選擇時的重要參考指標。

在以消費者為重的企業服務趨勢之下，民眾對於醫療服務品質的期望愈來愈高，而醫院亦愈重視病患滿意度，更成為醫療服務品質的重要指標之一。陳怡潔(1996)在其調查彰化某教會醫院，其門診病患對醫院的醫療服務品質的滿意度之結果發現，病患等候掛號時間、等候看診時間、等候批價時間、等候領藥時間愈長，則顧客滿意度呈現下降的趨勢。唯有醫師診療時間愈長，顧客滿意度呈上升的趨勢。可知等候時間對於病患滿意度之關聯性，可作為醫院服務品質不斷改善

的方向。

2.2 醫療品質指標與病患滿意度

醫療品質之指標可以從醫師、病患與服務滿意度來著手，而從結構、過程與結果等方面來加以評估。若從另一個角度來看，其組成維度包括醫院之資源分配、醫療品質與病患之特性，並藉由指標的異常回饋，以瞭解病患之期望與需求。在過去的醫療環境中，主要以醫護人員之醫療處置行為為主，民眾一旦生病就醫，醫院對醫療品質的定義僅侷限於醫療技術層面。隨著民眾對醫療知識之瞭解，加上消費者意識的抬頭，醫院間之競爭與經營成本壓力下，原本居於弱勢之病患，今已成為各醫院相互爭取之貴客。醫療品質不再等於醫療技術，為探討就診病患對醫療服務的滿意度，經由各學者的研究報告可以歸納出以下數項關鍵指標(Key Performance Indicators, KPI)，如表 2-1。由醫療服務品質面之 KPI 來看，各項指標中之等候時間為醫療服務品質的重要指標。

根據劉偉文(1999)研究探討屏東地區某教會醫院的門診病患對醫療服務品質滿意度調查分析，根據問卷調查的資料統計分析，得到的研究結果顯示在門診流程的等候滿意度方面。

- 現場掛號花費時間，在 20 分鐘內完成，是普遍可被接受的。
- 等候看診時間花費時間，以 5 分鐘以內滿意度最高，而 30 分鐘以上滿意度最差。
- 醫師看病花費時間，以 10 分鐘以內滿意度最高。
- 等候照 X 光花費時間，大約 30 分鐘內皆可完成，故整體滿意度頗佳。
- 等候檢驗花費時間，超過 21 分鐘以上者不滿意的程度相當高。
- 等候領藥花費時間，以 20 分鐘以內滿意度高。
- 等候批價花費時間，以 10 分鐘以內的滿意度最高。

表 2-1 醫療服務品質面之 KPI

1. 住院/手術病患滿意度
病患等候病床時間、辦理出院手續平均時間、供膳滿意度、住院整體滿意度、手術整體滿意度。

2. 門診病患滿意度
平均掛號時間、平均候診時間、平均領藥時間、平均檢驗/檢查等候時間、平均診療時間、平均批價時間、門診整體等候時間、門診整體滿意度。

3. 急診病患滿意度
平均掛號時間、平均候診時間、平均領藥時間、平均檢驗/檢查結果等候時間、平均診療時間、平均批價時間、平均會診等候時間、急診整體等候時間、急診整體滿意度。

4. 整體滿意度
整體病患滿意度、各項預防保健病患滿意度、內部員工滿意度。

5. 醫療訴訟糾紛
醫療糾紛比例、病患抱怨處理與比率。

其次在以某醫學中心之病患滿意度調查如表 2-2，門診病患認為該中心應該改進的地方：調查結果顯示，有二成門診病患希望改善「候診時間」(21.3%)與「環境設施及清潔」(20.4%)，一成希望改善「領藥時間過長」(13.8%)與「醫事人員服務態度」(12.8%)。(行政院國軍退除役官兵輔導委員會-統計出版品，2000)。

表 2-2 門診病患認為台北榮總應該改進的地方

(複選) (單位：%)

項目別	總計	榮民	非榮民
沒有需要改善的地方	17.8	31.5	14.1
醫事人員服務態度	12.8	16.1	11.9
候診時間	21.3	17.3	22.4
環境設施及清潔	20.4	7.7	23.9
領藥時間過長	13.8	11.3	14.4
掛號制度	8.3	8.6	8.2
其他原因	38.3	37.1	38.8
不知道/無意見	15.7	13.3	16.3

註：其他原因包含「行政人員服務態度」、「護理人員服務態度」、「看診時間」、「醫療技術」、「醫療設備」、「環境清潔人員態度」、「醫師資歷太淺」、「增加人手」、「行政效率不佳」等。

資料來源：行政院國軍退除役官兵輔導委員會-統計出版品 榮民總醫院就醫病患狀況調查報告
資料時間：88年 出版時間：89.12

2.3 醫院預約方式與制度

目前醫院在一般門診掛號方式，有現場掛號與預約掛號兩種。現場掛號主要以電話或親自到醫院掛當診看診之醫師與看診科別。預約掛號主要依科別與醫師指定非當診某段時間內到院看診，目前各醫院之預約掛號方式，隨著科技之發展已提供多元化之掛號系統，如電話語音自動掛號、網路掛號、觸控式掛號機及手機上網掛號等，而病患在此掛號系統下總是依醫院之預約制度，依醫師之特性予以等距法給號，若將現場掛號與預約掛號兩種交互穿插於就診序號時，則將會依現場掛號人次多寡而干擾預約掛號看診之預約時間。至於是否將現場掛號之病患安排於看診序號之前，或是與預約病患交互穿插於看診序號中，或者是安排於最後之序號，通常是依醫院對病患之整體考量而定。

2.3.1 醫院預約系統

目前醫院之預約方式，依預約之特性可分以下幾種：初診預約、複診預約、現場初診預約、現場複診預約與醫師約診等方式。不同的預約方式有不同設定給號方式，因病患之看診習慣與醫院之策略不同，加上醫院之資訊系統亦不相同，因此不同的醫院有不同的設定與操作方式。與此給號方式一般提供下列幾種資訊與使用方式：

- 可以在輸入畫面上顯示目前該診掛號，號碼使用的情況，此可提供預約給號考。
- 可以直接在畫面上點選給號，或電腦依各診規則以等距法自動給號。
- 可以個別指定及管控各診掛號起訖時間。
- 可以設定各診掛號限額。
- 各科可以設定各種不同規則給號，如單號為現場複診掛號，雙號為預約掛號。

表 2-3 為某醫學中心門診病患最常使用的掛號方式，其所佔比例分別為：「電話掛號」佔 51.4%、「現場掛號」及「人工預約掛號」佔 22.0% 為最常使用的掛號方式次序；榮民則以「人工掛號」（佔 59.2% 包含現場與人工預約）高於「電話掛號」（40.8%）。

表 2-3 台北榮總門診病患最常使用的掛號方式

項目別	單位：%		
	總計	榮民	非榮民
現場掛號	26.0	26.7	25.7
人工預約掛號（包含醫師診間預約）	22.0	32.4	19.2
電話掛號（包含語音、人工）	51.4	40.8	54.1
網路掛號	0.8	0.0	1.0

資料來源：行政院國軍退除役官兵輔導委員會-統計出版品 榮民總醫院就醫病患狀況調查報告
資料時間：88年 出版時間：89.12

2.3.2 醫院實施預約制度之優缺點

目前各醫院均有實施預約制度與現場掛號兩種方式，然而依各地方之民情不同，其實施結果亦大不相同。另外醫院層級不同，預約比例亦不相同。茲將其優缺點分述如下：

預約制度之優點：

- 預約制度可減輕等候掛號及候診室擁擠情況，使病人有更舒適的環境及提高醫院的服務水準。
- 提早調閱病歷，以降低現場病歷調閱時間，減少工作人員之壓力，與醫師等候病歷時間。
- 病患可有效掌控個人時間與時程的安排，不用擔心至現場掛不到號。
- 醫師可參考預約看診人數，調節可現場掛號人數，以控制合理門診量。

預約制度之缺點：

- 若掛號序號無法與預估看診時間結合，則會增加現場掛號之等候時間，病患容易流失。且病患趨向於就診前掛號，造成掛號之混亂與病患的抱怨。
- 需妥善設計預約掛號之看診序號方式，否則亦造成病患等候時間之增長病患的抱怨，如預約掛號看診序號排在前，對預約掛號比率較少之醫院，將流失現場掛號之病患。若現場與預約病患交互穿插於看診序號中，易造成兩種掛號病患之等候時間。

2.3.3 預約制度在各層級醫院實施方式與結果

由於各層級醫院之特性不同，其服務病患之本質亦將不同。從基層診所、地區醫院到區域醫院至醫學中心，病患看診基本上將區分輕重病而至不同之醫院看病，輕微小病可至診所或由家庭醫生診斷，而真正到需要更進一步之醫療設備或更精密儀器診療，才由診所轉診至專科醫院或大醫院做更進一步診療。現在很多

人認為大就是好之心理作用下，喜歡直接到大醫院專科去看病，造成輕重病混雜而影響醫療作業流程及醫師看診時間，預約制度之病患就診時間將難以掌控，所以造成病患等待時間長。至於基層診所，因其醫療作業流程簡單，加上地利之便，一般採預約掛號之情況較少，以現場掛號者居多。相對於大醫院基層醫師們期望的模式是不僅要建立轉診制度，各個層級醫院的功能也應劃分清楚，基層做些什麼、地區醫院做些什麼、醫學中心做些什麼皆有明確的內容，病人的問題在較上級的醫院獲得適當處置之後，可回到基層醫院繼續追蹤者即應轉診回基層醫院處置。理想的狀況是病人經基層醫師發現一些問題後，需轉診者由基層醫師與地區醫院預先做好聯繫，病人只要去到地區醫院就醫，也不用自己再掛號，再辦住院手續，其他一些後續的事情，我們這個體系都應該幫他做完善的處理。如此一來，醫療資源浪費的問題就可減少，病人也可獲得充分的照護。

依黃俊智(1992)之研究，在預約掛號與現場掛號之比率方面，各層級醫院亦有所不同。其中以醫學中心之預約比例最高，其次為區域醫院在次為地區醫院，基層診所預約掛號之比例最低。然而診所雖然有預約制度之系統，但卻沒有按照序號看診，病患若向診所預約看病，到了醫院並沒有享受到不用等候的方便，使得病患預約的意願降低。該個案診所如果推行依就診序號看診的方式，預約病患將可以減少等候時間約 600 秒，但現場掛號病患的候診時間卻會增加約 580 秒，使得整體病患的候診時間並無顯著差異。該研究進一步得之，適當的分散預約病患的到達時間將使得不論預約或現掛病患的候診時間減少。由於預約病患的預約到達時間間隔與醫師看診時間與醫師的預約人數有關，若醫師看診時間越長，則其預約時間間隔也應適度延長為宜，如果該時段該醫師的預約人數很多，則太長的預約時間間隔會使得預約病患會集中在看診時段後期出現，而醫師可能必須要延後下班，醫師在上班時段的實際看診時間比率也會因而下降，造成醫師時間的浪費。該研究結果顯示，若個案診所按照序號就診，且預約病患無遲、早到的行為時，則當醫師預約人數維持不變的狀況下，在非醫師甲看診的時段中，預約時間間隔改為 7 分鐘（原為 3 分鐘）可使病患總就醫時間減少 467.21 秒，預約時

間間隔改為 9 分鐘可使病患總就醫時間減少 536.40 秒；同樣的控制狀況而在醫師甲看診的時段中，則預約時間間隔改為 5 分鐘可使病患總就醫時間減少 324.53 秒。以個案診所目前的預約、現掛人數及預約間隔時間，假設診所依照序號大小看診，並且預約病患無遲、早到行為，若實施單號給預約病患、雙號給現掛病患方式，將可減少病患平均候診時間 307.55 秒。該研究雖然顯示依序號看診的方法與現況的差別不大，但由於預約制度可以使醫院提早進行病歷的調閱，減少臨時調閱工作的不確定性，並且良好的預約制度可以分散病患抵達醫院的時間，減少等候情形。

另外，侯幸雨(1998)之研究，在醫師閒置時間及病人等候時間，以模擬技術及等候理論之架構下有如下的發現：

- 在醫師閒置時間方面：
 1. 總病人數增加，相對地醫師閒置時間會減少。
 2. 爽約比率增加，醫生閒置時間則會增加。
 3. 預約比例增加，現掛比例較少的話，醫師閒置的時間也會較多。
- 在病人等候時間方面：
 1. 病人爽約比率越多，等候時間就會減少。
 2. 預約比例越高，現掛比例較少時，等候時間也會減少。
 3. 至於病人總數多寡，在預約病人六成、現掛病人四成、總病人數為 45 人時，等候時間為最長；其他在預約四成或五成時，病人總數 30 人之等候時間最長；其與醫師服務時間之長短，及預約病人到達時間之間隔有關係，所以呈現出不同的結果。

2.3.4 門診等候時間各屬性之資料分析

在門診看診流程之等候時間方面，盧昭文(2000)以某區域醫院之資料來探討門診流程，從病患開始掛號，看診，批價到領藥時間之特性關係。每位門診病患在醫院的正常作業流程下，平均排隊等候掛號時間花費 4.5 分鐘，在候診區等候

醫師看診時間花費 15.1 分鐘，進入診間就診時，醫師看診時間平均為 6.2 分鐘，排隊等候批價時間為 4.5 分鐘，最後在藥局前，排隊等候領藥時間花了 2.7 分鐘，以上不含放射檢查與檢驗檢查所需時間。從病患到院至離院，完成整個看診作業流程，平均總共需花費 33.1 分鐘。如表 2-4

表 2-4 門診看診流程之等候時間分析表

項目	樣本數 (人)	平均值 (分)	最小值 (分)	最大值 (分)	櫃檯數 (個)	平均累積 等候人數
等候掛號時間	231	4.5	0	14	4.8	23.3
等候看診時間	2262	15.1	0	99	—	—
醫師看診時間	2262	6.2	1	51	—	—
等候批價時間	227	4.5	0	14	4.9	23.5
等候領藥時間	128	2.7	0	10	2	—

由上述資料可知，整個看診之流程路徑，從掛號、等候看診到批價最後到領藥，其中以等候看診時間為最長。約佔整個看診流程 56% 之等待時間(醫師看診時間除外)。

2.4 資料探勘

所謂知識發掘(Knowledge Discovery)就是從大量資料中，尋找其規律性及其中隱藏的知識或未知的事實，希望能以此來作為決策支援之有力輔助工具。目前各大醫院每日累計了大量門診、住院、檢驗、檢查等重要病患資料，但是資訊多樣性與無結構化之特性，使得醫院決策單位護理人員或醫師，無法有效利用現存的資料轉換成有效之資訊，甚至因無法即時獲得有效資訊，使決策行為產生重大損失或錯誤。如果能讓病患或醫護人員透過電腦，即時且精確的提供個人化病

歷資料，透過知識之表現與規則之建立，以資料探勘(Data mining) 和資料倉儲 (Data Warehouse) 之密切整合之技術，解決大量病患歷年資料儲存之空間問題，及即時查詢之效率問題，以期建立個人化專屬健康資訊與病歷系統。讓醫師能即時得到病患歷年之疾病趨勢或血壓及相關生理資訊演變圖，另一方面以資料探勘提供同一疾病之相關醫療決策支援資訊，以了解社區之疾病發生與個體之病患之關係，而達到真正之社區醫療健康照顧，並提昇醫院專業與病患之滿意度。

以知識庫為基礎之醫療健康管理，其重點在於要結合醫療領域方面的專業知識，及建立完善電子病歷之資料倉儲，才能有效將資料(Data)轉成資訊(Information)進而轉成企業的知識(Knowledge)。換句話說，為獲得專業之醫療品質，醫院必須以一對一之個人化病歷，加上族群化之病患分析與決策資訊，來分析個人之健康與族群間疾病關係，達到預防醫學之目標。

2.4.1 Data Mining 的功能

Data Mining 先驅 Jonattan J. Palmer (1997) 認為建置 Data Mining 最好的方法是分期分步驟進行，以遞增式的方法開發。他需要一些資料的前處理以及選擇合適的演算法。以下就 Data Mining 的功能、建置方法以及計畫的擬定等作進一步的介紹。

一般而言，Data Mining 技術可達到下列五項功能：分類(classification)，推估(estimation)，預測(prediction)，關聯分組(affinity grouping)以及叢集或稱同質分組(clustering)。茲將這些功能的意義及可能使用的技巧簡述如下。

- 分類

按照分析對象的屬性分門別類加以定義，建立類組(class)。例如，將信用申請者的風險屬性，區分為高度風險申請者，中度風險申請者及低度風險申請者。使用的技巧有決策樹(decision tree)，記憶基礎推理(memory - based reasoning)等。

- 推理

根據既有連續性數值之相關屬性資料，以獲致某一屬性未知之值。例如按照信用申請者之教育程度、行為別來推估其信用卡消費量。使用的技巧包括統計方法上之相關分析、迴歸分析及類神經網路方法。

- 預測

根據對象屬性之過去觀察值來推估該屬性未來之值。例如由顧客過去之刷卡消費量預測其未來之刷卡消費量。使用的技巧包括迴歸分析、時間數列分析及類神經網路方法。

- 關聯分組

同一個交易中，一個物件出現也會引起另一個物件的出現，從所有物件決定那些相關物件應該放在一起。在客戶行銷系統上，此種功能係用來確認交叉銷售(cross selling)的機會以設計出吸引人的產品群組。例如，超市中相關之盥洗用品(牙刷、牙膏、牙線)，放在同一間貨架上。一個有趣的關聯例子：以家樂福量販店對交易資料庫作探勘，可發現如下之規則。

對任何一筆交易 T_i ，資料集 D 為購買物品(d_1, d_2, \dots, d_j)

$T_1 : (d_{11}, d_{12}, \text{麵包}, \dots, \text{牛奶} \dots)$

$T_2 : (d_{07}, \dots, \text{麵包}, \dots)$

$T_3 : (d_{33}, \dots, \text{牛奶}, \text{麵包}, \dots)$

$T_4 : (d_{22}, d_{34}, \dots)$

$T_i : (\dots)$

此表示在超商購買麵包時，他很可能也會購買牛奶。至於有多高的可能?可以關聯規則之可信度(confidence)加以表示。可信度為：在 A 出現之條件下出現 B 之機率。

可信度 = $P(B|A) = P(A,B)/P(A)$

$$P(\text{牛奶}|\text{麵包}) = \frac{P(\text{麵包}, \text{牛奶})}{P(\text{麵包})}$$

- 叢集

將異質母體中區隔為較具同質性之族群(clusters)。其主要之特性為群內同質；群間異質，但是，假定事先未對於區隔加以定義，而資料中自然產生區隔。使用的技巧包括如 k-means 法，EM(Expectation-Maximization)及 Agglomeration 法，而 EM 為本研究使用之主要方法。

2.4.2 資料探勘(Data Mining)建置步驟

在此階段依建置的先後次序加以說明，此建置一共包含了六個步驟分別包括資料選擇、過濾與清除、充實資料、編碼與資料轉換、資料探勘與資訊報告(Tsai, Yuh-Show, 1997, 吳國禎, 2000)。其流程如圖 2-1。

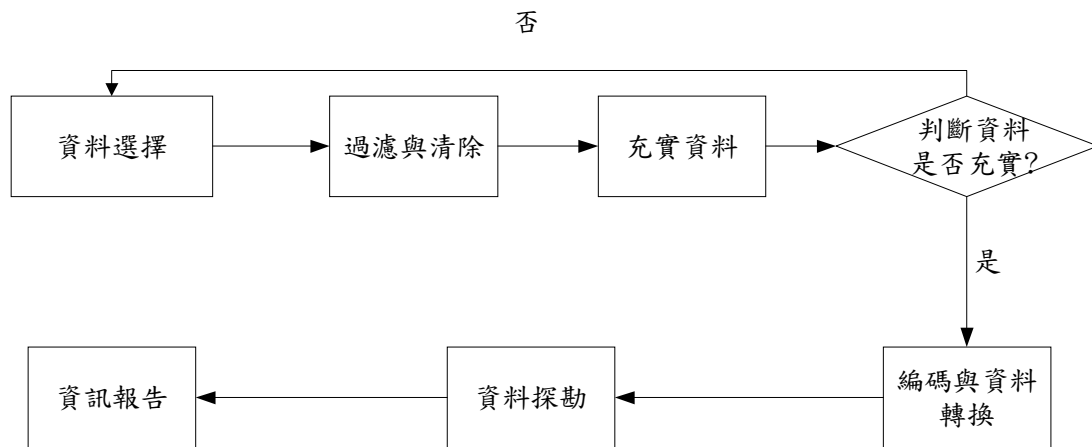


圖 2-1 資料探勘建置步驟

1. 資料選擇：在此步驟，從事使用者需求訪談，以充分了解該應用領域之特性，其目的在於熟悉相關的先導知識後，可了解使用者的目標為何，以便正確選擇所需資料。
2. 過濾與清除：一般之原始資料常有無法預期之錯誤，或含有重複性及不一致之資料，例如年齡為負值，或已有疾病碼又有中文及英文疾病名稱，此時可選擇疾病碼而中文及英文疾病名稱當參考即可。另外在資料一致性上，多重

運作系統中相同資料用不同方式來表現時是可能發生的。例如：在一個運作系統中一個藥品名稱會被縮寫，但在另一個運作系統中藥品名稱卻沒有縮寫。假如沒有使兩個值達成一致性，任何的查詢使用到該資料時有可能會計算出不同值的產品。

3. 充實資料：此部份一般在執行資料探勘時，可依所需資料之狀況及資料探勘結果加以補充或是減少資料以便了解不同結果之差異。
4. 編碼與資料轉換：由於有些欄位是由原始之方式來表現，其資料較為複雜且不易於資料探勘使用，此時需將資料加以重新編碼及轉換。例如病患看診日期一般表示為 20020526 此代表 2002 年 5 月 26 日，然而為了探討病患看診星期與看診人數的關係時，必須將日期轉換為星期，如 2002 年 5 月 26 日以文字型態 7 來表示該天為星期日。
5. 資料探勘：選擇適當的演算法及其對應之參數設定，將經過前述步驟的資料，經由訓練與學習過程後，找出隱藏的資訊，並依照不同的演算法與該領域之瞭解，以建立一可用的模式。
6. 資訊報告：當一模式建立完成後，可利用這些模式的特徵(Patterns)以及關聯性(Relationship)以提供決策所需資訊或預測事件的發生。

2.4.3 資料探勘與資料庫倉儲

資料倉儲系統是一種能適當對資料做整合及管理不同資料來源的技術，強調的是資訊 (Information)，而非資料 (Data) 的分享。此整合性資料儲存體，能提供企業解決問題及決策輔助，可供高階主管做相關資訊查詢、擷取、篩選、整合。資料倉儲所欲解決的問題是主管決策分析的問題，而非一般線上交易導向的問題。對於資料擷取方面其不同於傳統系統的被動式 SQL(Structured Query Language)查詢，而資料倉儲系統為主動資料導向式的查詢，當來源資料更動時即做出相對應的反應(謝崇耀，2000)。為了提高資料的查詢速度，資料倉儲則根

據不同之需求，建立了各種維度(Dimensions)之資訊，因而產生了一個非常大的資料庫，它儲存著由各部門及客戶作業資料庫中整合而來的不同資料，特別是從線上即時應用處理 OLAP (Online Analytical Processing)所得的資料與從分散異質資料資源傳遞而來的資料，因此將作業層次的資料轉換成有用的策略性資訊，是整個資料倉儲系統建置的重點所在(Golfarelli & Rizzi，1999)。

圖 2-2 一個典型資料倉儲的組態(Connolly，1998)，資料來源是多樣的，資料先從來源系統(Source)經由載入管理系統(Load manager)移到載入區(Meta Data)，接著在載入區進行轉換將資料依所需分別載入細部資料區、中度資料整合與高度資料整合區。一旦資料載入到資料倉儲，使用者將可透過查詢管理(Query Manager)，隨心所欲地使用主管決策工具 EIS(Executive Information System)、線上即時應用處理 OLAP 或進行資料探勘。

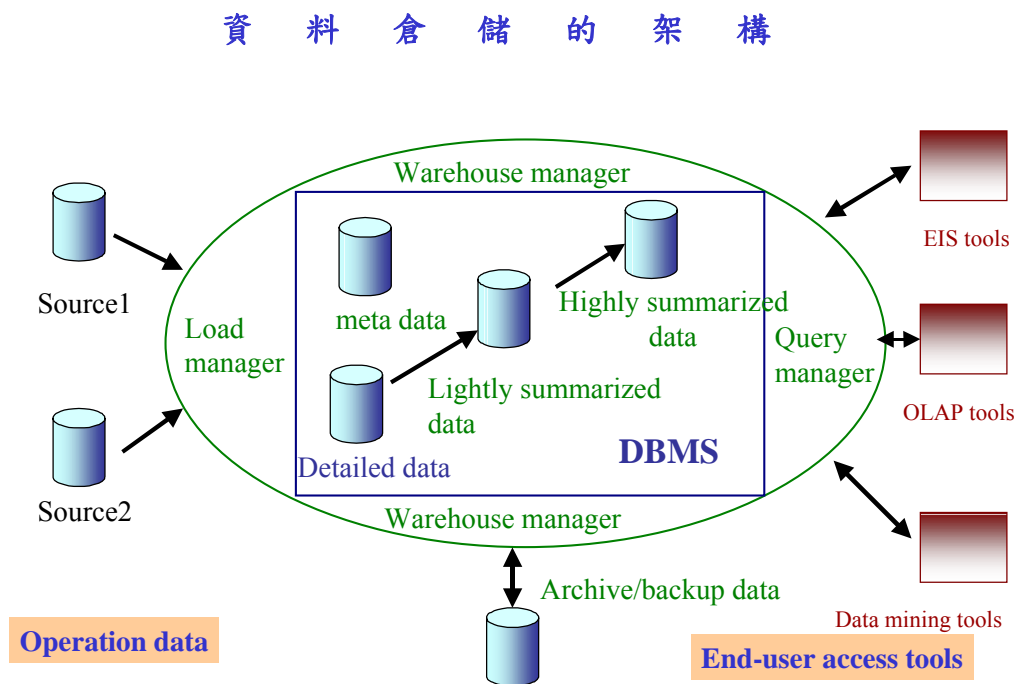


圖 2-2 一個典型資料倉儲的架構

「資料探勘是從一個龐大的資料庫中將正確、以前未發覺卻非常重要的資訊加以抽離出來，並利用這些資訊來做出重要決策的過程。」(DB2 資訊月刊, Peter

Gwynne, 1996)。由上述的定義可以很清楚地知道，資料探勘強調的是將眾多之資訊(Information)轉換成以前未發覺之知識(Knowledge)。資料探勘已不是依已知條件，結構化對資料庫提出一個查詢的要求而已，事實上，它是一個更加複雜的程序；利用它，使用者可以從無結構化及排列方式雜亂無序之資料中，找出並規劃出有用的資訊。此有別於資料倉儲以預先知道之方式，建立多維度資料查詢方式相比較，在做資料倉儲查詢時，原始的資料必須以某種特定的方式安排，例如要查詢醫院病患以語音、電話或現場掛號方式所佔比率與人數，則以特定的時間維度將上述資料加以排列整合，當要以週、月或年為資料維度時，即可以立即查詢。而資料探勘則將大量之資料依不同之演算法，建立出特定之知識模式，例如以非監督式(Unsupervised Learning)的知識發現技術，可以不需事前鎖定探索主題與資料類別下的情況下，依資料之邏輯特性，找出以前未發覺卻非常重要的資訊。

2.5 等候心理學

在 David Maister 之服務管理中提出兩項服務法則：其一是有關顧客期望與認知的比較，例如當病患在所就診之醫療院所下，所接受到的服務超乎期望，便會快樂，並將這項好的服務告知友人，使該醫院服務因口耳相傳效果而受益；然而，口耳相傳效果也可能使服務不好之醫院，而得到不好的名聲。其二是顧客的第一印象會影響剩餘的服務經驗，因此需讓病患等候的服務應注意，使等候期間成為一種愉快的經驗；對病患來說，等候的認知通常比真正花在等候的時間更重要，所以服務提供者必須考量下列幾種等候心理(Fitzsimmons, 2002)：

1. 空白的時間

人們不喜歡空白的時間，因其不但使人們無法從事身心的活動，而且帶來不好的感覺，尤其是它似乎是永遠持續的，因此醫院必須以正面的方式填補這段時間。例如：如何將無聊的等候時間變成是有趣、有意義的活動，即可轉變病患的心理態度。而改變之方式可如下之範例：

- 可在候診區提供電視以轉移病患等待的心理。

- 可在候診區增添報紙刊物以供閱讀或加貼衛教海報。
- 候診室外較大的活動空間可固定安排美展或兒童園地。

2. 轉移行為

在等候之過程中，有些轉移行為祇是為了使病患不會感覺等候很久，有些則還可以為醫院因服務性質之特性，而帶來額外的收入，例如醫院附設藥局或書局，可使病患於等待時間上，可同時購買相關醫藥產品或有其他之活動。此外，與服務有關的轉移行為可以傳遞服務已經開始的訊息，因此降低病患的焦慮程度。例如病患到院時，量血壓或體重之服務。與服務尚未開始相比較，當病患感覺服務已經開始，便可以容忍較長的等候；而且病患對起初的等候比服務已經開始的等候更容易變得不滿意。

3. 未知的焦慮

在服務開始之前，病患可能存有許多未知的焦慮，這些正是影響等候時間長短的心理因素。因此醫院必須發展緩和焦慮的策略。例如：等候領藥或就診時有號碼指示燈號，以告訴病患大概要等多久。

4. 不確定性的遊戲規則

不確定性與未解釋的等候會造成病患的焦慮，而當病患看到晚來的人先接受看診服務時，焦慮會轉為對不公平的憤怒，而醫院之服務人員與篡位之病患均會成為憤怒的目標。為避免違反先到先服務（First-Come First-Served, FCFS）的等候規則，可以採用「取號碼牌」的策略，以降低顧客對等候長度與可能之不公平性的焦慮。此外，當有多個服務員時，採用「取號碼牌」的單一等候線策略以維持 FCFS 規則。例如醫院之掛號櫃檯，因為病患客可以確保取號碼牌公平性，所以可以放鬆心情觀看附近之衛教宣導，如此不但佔據了病患的空白時間，而且可以使等候時間感覺較短，並提高醫病關係之相互溝通。但並非所有服務都有明確的優先順序，例如急診服務需依檢傷分類病患之嚴重度給予優先處理，不過為避免激怒其他病患，造成明顯的歧視印象，最好能明顯告知這類特殊的處理，以避免病患之抱怨。

5. 社會正義的程度

Richard Larson(1987)觀察到在等候滿意度中關鍵的決定因素是社會正義的程度，即使等候時間相當短，若是服務系統違反先到先服務的原則，顧客也會被激怒。

最後，David Maister(1984)提出 8 個等候理論之原則如下：

- 空閒的時間感覺比忙碌的時間長。
- 過程前的等待感覺比過程進行中的等待來得長。
- 不安、渴望使等待似乎更長。
- 不確定性的等候比已知固定的等候更長。
- 沒有解釋的等候比解釋過的等候長。
- 不公平的等候比公平的等候長。
- 服務愈有價值，人們愈願意等候。
- 獨自等候的時間長於群體等候。

2.6 病患等候時間與相關改善研究

在 1950 至 1980 年間門診病患預約看診時間，在國外已有廣泛的研究(Blanco White 1964，Vissers 1979)。然而因國外之健康保險制度與本國不同，就醫型態亦大不相同。在 University of Iowa, 其病患預約間隔時間為 34 min(Dexter, 1999)，在台灣各大小醫院平均間隔約為 3 至 8 分鐘，因國外之相關病患等候研究將與本國大不相同，其實用性亦不適用於本國，茲就本國相關之研究重點分述如下：

利用等候理論與電腦模擬技術為研究工具，來探討醫院的門診看診服務情況，並結合目前普遍之行動科技理念如 PDA、手機等，以更彈性之方式告知病患看診時間(古政元，2001)。

在對病患等候領藥作業流程加以改善方面，根據成大醫院資料統計結果得

知，藥品項數較少者佔所有調劑單的大多數，故模式將根據其現有的設備和收集的資料，將調劑單依藥品項數多寡分為單號和雙號兩條等候線，取代舊有隨機分配的等候線，以分析此系統在等候時間最小下，該以品項數多寡來分界，並同時考慮在不同日期、不同時段下，各個服務台之最佳藥師數配置，做為醫院管理者的決策指標，以期滿足醫院的經營理念和減少病患的等待領藥時間（吳永順，1998）。

在醫院掛號作業效率之決策研究中，也有引進生產業的「作業研究」（Operations Research, OR）理論，以解決等候時間最佳的方式。此模式經檢定後，證實到達率與服務率均屬指數分配，經模式設計計算出等候理論的四個值 L 、 Lq 、 W 、 Wq 。其代表意義如下：

L ：在系統中的期望顧客數。

Lq ：在等候線上的期望顧客數。

W ：每位顧客在系統中的期望等候時間。

Wq ：每位顧客在等候線上的期望等候時間。

再以去年同期的業務量為基礎，乘以近三個月的平均成長率及時段比率作調整，來預測下個週期的業務量（到達率）。接著以抱怨等候時間過久而影響下次再來院看診的損失為等候成本，加上人員成本，來選擇最低總成本的組合，找出最佳人力排程。經實證結果，以等候成本比較節省 8101 元/週(32.5%)，在病患等候時間縮短在可接受範圍，服務人員忙逸度亦較平均，尤其是顧客滿意度提高，實證效果良好，應具實用價值(楊朝欽，1996)。

在醫院門診掛號時間之研究，利用等候理論及類神經網路中的倒傳遞網路做為研究工具，以設計醫院的掛號預約系統，用此設計出的掛號預約系統，能讓病人平均等候時間縮短在 20 分鐘以內，且病人等候超過 40 分鐘的人數相對的減少(胡國岱，1996)。

第三章 叢集

3.1 叢集之基本概念

本研究將對叢集使用的基本概念以及對本研究有關之理論加以介紹。另外對於叢集之分類法以及一些重要的叢集概念亦加以簡要的說明。

當我們遇到一個具有高維度且複雜之大量資料時，最直接有效的方式就是將資料依其相似性加以分組。目前叢集已廣泛地應用於各研究及實際應用領域，例如群組分類、文件擷取及影像分析等，在此將說明叢集之基本特性與應用。基本上叢集是將資料加以分類成群，同群間資料相似性高，群與群間其資料異質性高。以叢集之方法論，問題可以在不需事前鎖定探索主題的情況下，而適用於不同形式的資料，例如以文字及數值型之各種變數資料型態，他是屬於非監督式(Unsupervised)的知識探勘技術。然而依問題探討之模式不同，應用資料探勘之方法亦大不相同，在此說明所謂非監督式(Unsupervised)與監督式(Supervised)之不同。以監督式分類而言，一般都先得知已分組(Labeled or pre-classified)之模組，再依此模組去歸類未編組之資料，簡而言之，就是規則的產生是依已知的模式。而所謂的非監督式是指對於資料之型態事先並未給予先分類，而將此未分組(Unlabeled)之資料依其自然之分布特性歸類於某一群組，以顯示某些相似的、可以預測的特徵。通常，這些特徵會被隱藏或不容易被察覺。

以下為叢集的一個簡單例子如圖 3-1：

由圖 3-1 之例子我們可以了解，身高與體重可區分出不同的兩大族群，圖中分別以紅色族群與綠色族群表示，此兩族群之特性而言，性別並沒有任何區別。對同一族群其相似性高，而不同族群間其相異性高。因此對於群集化並沒有依靠事先明確定義的類別來進行分類，而且群集的意義要靠事後的闡釋才能得知。

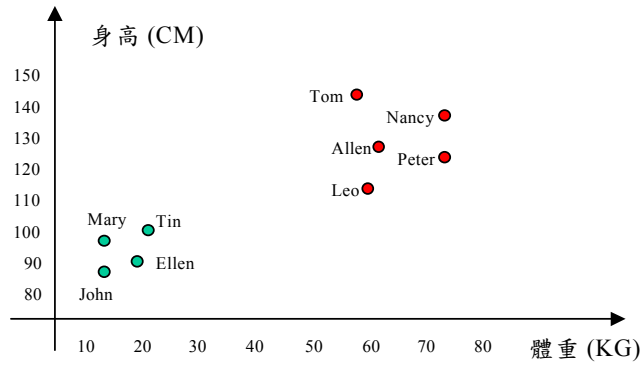


圖 3-1 對於身高與體重與性別叢集的結果

3.2 叢集建置之階段工作

在叢集趨勢(Cluster tendency)之研究中，主要在探討在執行叢集前，對於輸入的資料先做一檢驗以確認是否有相關，為了使叢集建置有效進行，基本工作上包括下列幾個階段(Jain, 1988)：

3.2.1 模式表現 (Pattern representation)

模式表現通常包括選擇叢集的數目、可使用的叢集模式以及叢集演算法中可使用的資料數量、型態及其範圍大小。其他之選項包括：特徵選擇與篩選(Feature selection and extraction)。特徵(Feature)在此之定義為：描述一個物件之屬性與其所代表的值。例如：眼睛之顏色為一屬性，黑色的眼睛則為一特徵。

所謂特徵選擇：指的是在叢集之所有特徵中，如何選取最有效率之子集。為什麼要做特徵選擇？因為太多的特徵會使我們分析時花費很多時間，另外一些不相關之特徵，因給予事先清理，因此選擇適當之子集為選擇特徵之重點。

特徵篩選：是指在輸入之特徵中，如何利用一個或多個轉換去產生更顯著之特徵。不論特徵選擇或特徵篩選以上兩種技術均可適當的得到叢集所需之特徵。而特徵選擇之方法主要分為逐步式選擇與嘗試錯誤型之特徵選擇兩種方法：

1: 逐步式選擇：

逐步式選擇又可區分以下兩種：

- 向前選擇法(Forward selection)：選擇認為適當之特徵，加入子集以了解叢集之結果與效率，並選擇那些可增加系統效率之特徵。
- 向後消除法(Backward elimination)：從現有之特徵群組中，逐步刪除一些特徵，該特徵並不會造成系統及效率之影響者。

2: 嘗試錯誤型之特徵選擇：

嘗試錯誤型之特徵選擇，主要依下列準則選擇最少之特徵子集：

- 分類結果的精度並沒有明顯的降低。
- 經由選擇後的結果分類，盡可能的符合原始分類的分佈。

唯一能滿足以上兩個準則之條件就是仔細的尋找可能的特徵。

3.2.2 相近模式(Pattern proximity)

指定義模式間之距離量測方式，其用以尋找與目標物件(target sequence)向量距離越短越好的物件，其表示式如下：

$$\min \sum_{i=1}^n (X_i - K_j Z_i^J)^2$$

以下則為不同之距離量測方式表示法：

- Euclidean Distance (Pythagorean distance, Grower, 1971)

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ip} - x_{jp})^2}$$

- Manhattan Distance (city block) (Carmichael and Sneath, 1969)

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \cdots + |x_{ip} - x_{jp}|$$

- Minkowski Distance :

$$d(i, j) = (|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \cdots + |x_{ip} - x_{jp}|^q)^{\frac{1}{q}}$$

- Weighted Euclidean Distance :

$$d(i, j) = \sqrt{w_1(x_{i1} - x_{j1})^2 + w_2(x_{i2} - x_{j2})^2 + \cdots + w_p(x_{ip} - x_{jp})^2}$$

不同的距離量測法適用於不同之叢集分類法，例如：以一個最簡單的距離量測法來說，歐幾里德距離(Euclidean Distance)，通常可用來表現兩個模式間之相異性。距離越遠，其相異性越高；反之距離越近，其相似性越高。

3.2.3 叢集分類(Clustering taxonomy)

叢集之分類可有不同的實施方法，依其實施方法不同，輸出之叢集可為明確之資料族群，或為部分重疊。例如，階層式的叢集演算法依資料屬性的相似性程度，將資料進行逐次合併為一系列的階層式樹狀架構。分割式的叢集演算法是從大量的資料中區分所有可能的相似叢集。以 K 平均法(K- Means)而言，將每一資料點分配到質心最接近的群集中，再計算每一個群集的質心，再將群集中每一個點的位置加以平均找出新質心，每一點再次被分配到質心最接近的群集中，重複進行直到群集邊界不再變動為止。

3.2.4 評估叢集的結果(Assessment of output)

以各種不同的叢集演算法而言，可獲得不同相似程度的叢集，對資料集可透過叢集運算的結果，區分資料間相似與相異性。對不同資料而言不同方法可能有不同結果。但當我們有了這些結果後，如何去評估一個叢集結果的好壞？將直接影響決策的品質與成敗。至於要如何評估其好壞，可參考以下之建議方式：

- 依資料特性研判其呈現結果，而不要完全依賴叢集演算法本身的結果而下結論。採用此方法亦須了解本身研究領域的專業知識，才可對叢集的結果做一合理的解釋，對於資料本身愈了解愈能對群集之結果做更正確的評斷。另外對找出來的群集並無法保證擁有任何實用價值。
- 資料如果確認樣本數太少，不構成代表性，或資料單一性太高，無群組的特性，其結果將無法合理解釋。就不要用叢集演算法。

3.3 叢集技術的分類 (A Taxonomy of Clustering Approaches)

叢集對於資料的處理，有不同的方法，不同觀點亦有不同的歸類法，圖 3-2 為 Jain 及在 1988 年將叢集的方法以樹狀圖分類加以表示(Anil & Dubes, 1988)。

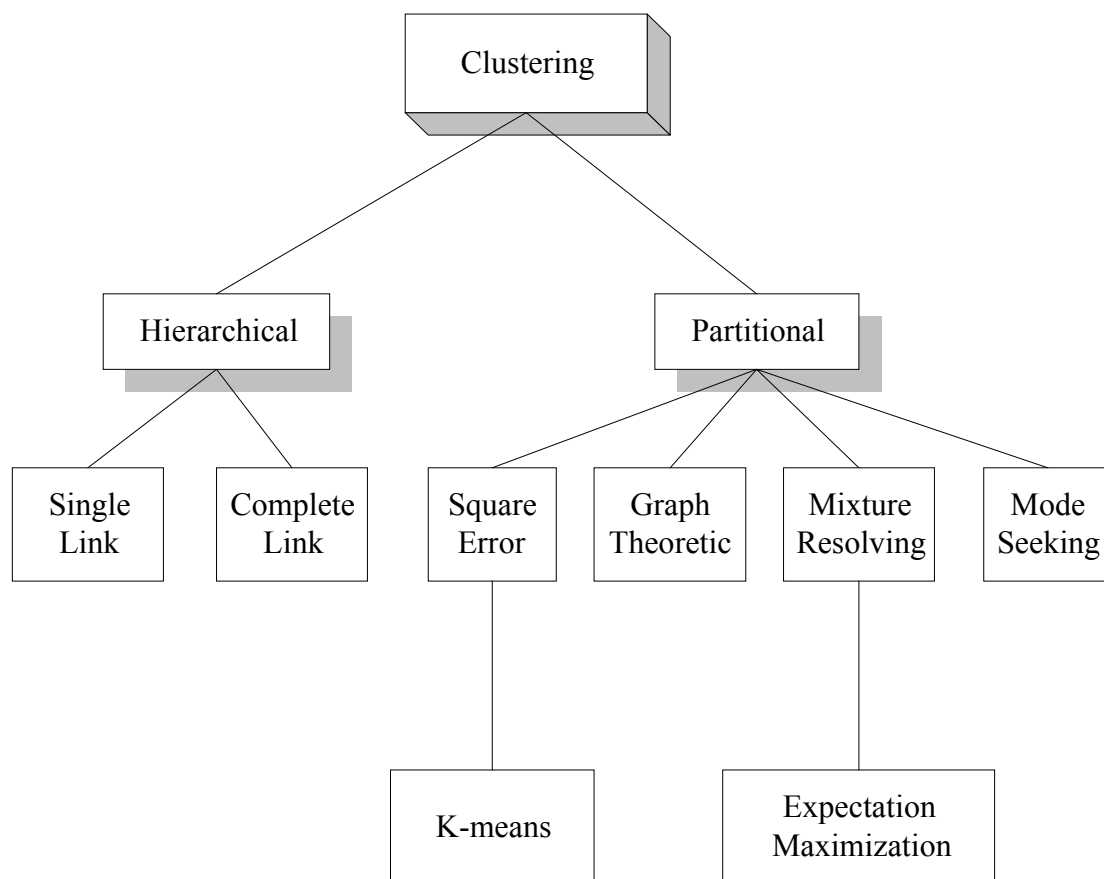


圖 3-2 叢集技術的分類

圖 3-2 為叢集技術的分類，此樹狀圖之分類架構而言，主要區分為兩大類，一為階層式(Hierarchical)另一為分割式(Partitional)。階層式則分為單一連結(Single Link)與完整連結(Complete Link)，其主要特徵為利用凝聚演算法(Agglomeration Algorithms)，將一個個單一的叢集(A distinct singleton cluster)合併，直到滿足此準則為止，其結果將產生一系列樹狀式之分割。分割式(Partitional)的主要方法為 K-means 及 EM(Expectation Maximization)。K-mean 主要是應用反覆式距離為基礎的叢集法則，EM 主要是應用可能性分佈叢集法則(Probability-based clustering)。

3.4 叢集的演算法

依據圖 3-2 叢集技術的分類，以下將介紹常用的幾種叢集演算法，包括：階層式叢集演算法，分割式叢集演算法、平方差演算法、K 平均叢集演算法及 EM

演算法。

3.4.1 階層式叢集演算法(Hierarchical Clustering Algorithms)

階層式叢集演算法，其架構為一層級式群集，每一層級的改變代表群集的改變。其主要的演算法包括以下三種：

- 單一連結法 (Single-link Clustering)。
- 完整連結法 (Complete-linkage Clustering)。
- 比較質心法 (Comparison of centroids Clustering)。

為便於了解各個演算法，以下將舉一例子以圖例方式簡單加以說明，何謂階層式叢集演算法中之單一連結法。

單一連結法

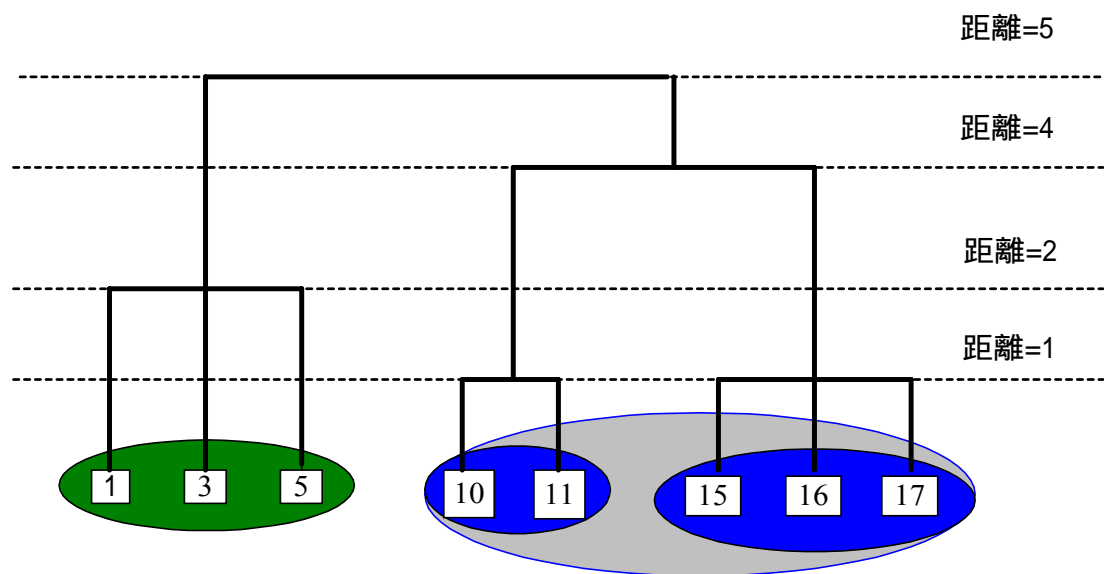


圖 3-3 單一連結法凝聚步驟之圖例

圖 3-3 將舉例說明單一連結法之凝聚步驟：

1. 首先將資料以相似程度作階層式排列，如圖將數字從 1, 3, 5...排至 17。
2. 從排序的資料中，找出兩兩距離 Δd 之值，在將相同 Δd 之值凝聚成一小

群集。如圖之(1、3、5)，(10、11)，(15、16、17)。

3. 最後的樹狀式架構，則是將兩兩之小群集，取其最小的距離而得到的結果。而在完整連結法，則是取其最大的距離而得。比較質心法則是取各小群集質心距離而得。

3.4.2 分割式叢集演算法(Partitional Algorithms)

分割式叢集演算法，是從大量的資料中區分所有可能的相似叢集，此與階層式加以比較，此方法的優點在於可處理大量的資料，而此正是階層式方法無法有效加以計算獲得的。然而分割式叢集演算法伴隨著的問題是，如何選擇最佳輸出叢集的數目，在此方面可從學者的研究中，提供一些指導原則(Dubes 1987)，雖然此原則有助於叢集的數目的選擇，而在實際的應用中，還是要以不同之狀況，依叢集輸出的結果，做最後的判定。

以下將說明分割式之常用演算法，包括：平方差演算法及 K 平均叢集演算法。

- 平方差演算法 (Squared Error Algorithms)

在分割式叢集的技術中，最常使用的準則為平方差演算法，以下為其定義之說明，對一個包含 k 叢集模式組的叢集(\mathcal{L})其平方差演算法如下式：

$$e^2(\mathcal{L}) = \sum_{j=1}^k \sum_{i=1}^{n_j} \left\| x_i^{(j)} - c_j \right\|^2$$

在此 $x_i^{(j)}$ 是 i^{th} 模式(pattern)隸屬於 j^{th} 之叢集， c_j 是 j^{th} 之叢集之質心。

以下為平方差叢集法的導入步驟：

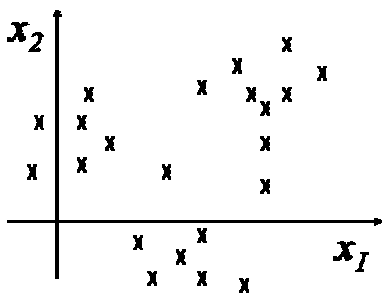
1. 在一資料集當中選擇初始模式的分割 (partition of the patterns)，而此模式的分割具有固定的叢集數及其質心。

2. 將每一資料點分配到質心最接近的群集中，計算每一個群集的質心，將群集中每一個點的位置加以平均找出新群集之質心，每一點再次被分配到質心最接近的群集中。重複進行直到群集邊界不再變動為止。
3. 依據經驗法則將群集合併或分開，在不同的叢集數下，從新執行第二項步驟，以得到合理的模式結果。

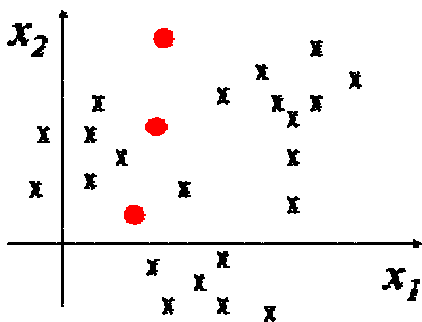
• **K 平均叢集演算法 (K- means Clustering Algorithm)**

K 平均法是最簡單且普遍應用平方差的一項演算法，以下為其方法的執行步驟與簡單的圖例：

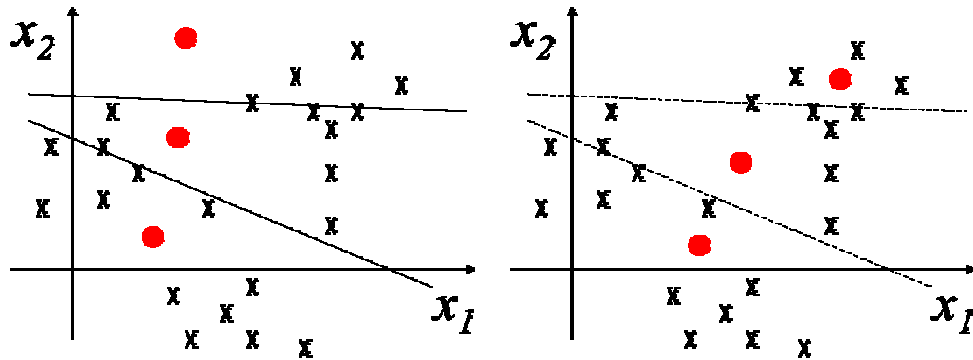
1. 假設有一組特徵向量以座標 x_1 及 x_2 來表示



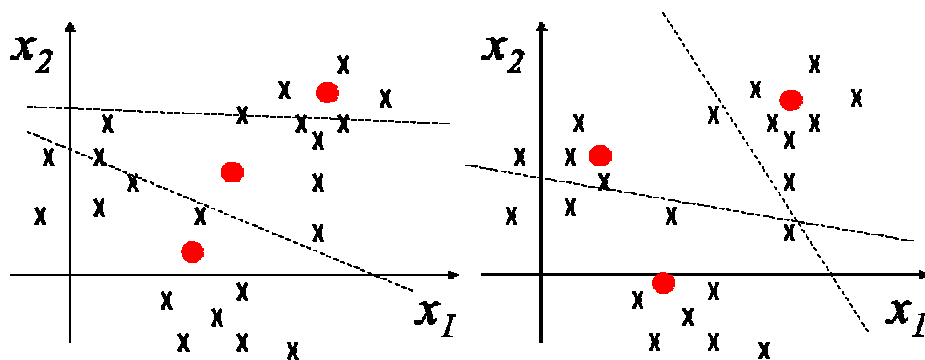
2. 任意選擇 K 個「種子」悉按亂數設定作為群集質心，假定 $K = 3$ 如圖中之三小紅點。



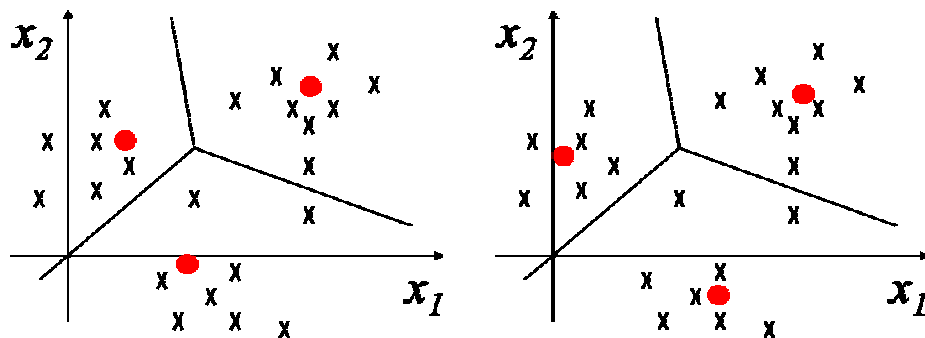
3. 將每一資料點分配到質心最接近的群集中，計算每一個群集的質心，左圖為計算質心前，右圖為計算質心後。



4. 將群集中每一個點的位置加以平均，找出新群集，每一點再次被分配到質心最接近的群集中。左圖為計算質心前，右圖為計算質心後。

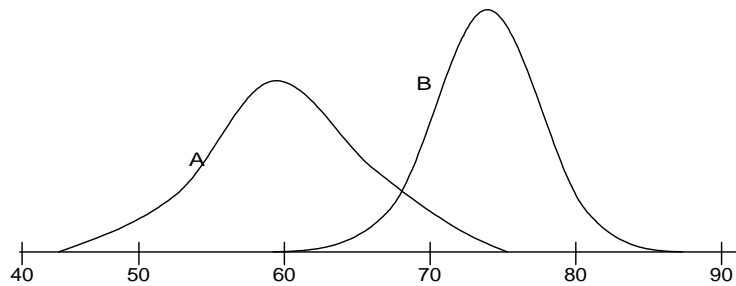


5. 重複進行直到群集邊界不再變動為止，右圖為最後之結果。



- 機率基礎的叢集(Probability-based clustering)

此章節將說明一個廣泛使用於無法直接觀察到變數值的學習方法，從機率觀點來看，此方法的目的將從已知的資料集中，找出性質相似的資料群組，在其他的分類中亦稱為機率基礎的叢集(Probability-based clustering)(Lan, 1999)。在此資料集中，每一個資料將依其特定之機率與特性而有其個別的分佈，而被歸類於某一特定的叢集。對此資料集而言，若有 K 個機率分佈，我們就稱為有 K 個叢集，而對產生的每一叢集其所佔的比率在整個母體而言並不一定會相同。在說明 EM 之前，先以一個簡單 Finite Mixture 來加以說明，此例中只有一個數字型態的屬性(Numeric attribute)，且每一個族群皆具有高斯(Gaussian)或常態分配，但不同的平均值 μ (Mean)及標準偏差 σ (Standard deviation)。此叢集的問題為：如何從這些已定義出叢集數目的資料中找出各個叢集的平均值及標準偏差，以及其在整個母體中其所佔比率之多寡。



$$\mu_A = 60, \sigma_A = 5, \rho_A = 0.6 \quad \mu_B = 75, \sigma_B = 3, \rho_B = 0.4$$

圖 3-4 為 A 和 B 兩常態分配圖及其分佈比率

圖 3-4 為一簡單例子的說明，在圖中有 A 和 B 兩個叢集，其平均值及標準偏差分別以 μ_A 及 σ_A 表示 A 叢集，以 μ_B 及 σ_B 表示 B 叢集，機率分別以 P_A 及 P_B 來表示($P_A + P_B = 1$)。剛開始我們假想只有資料集及族群數目而沒有已分好的族群，現在我們將決定如何獲取上述五個參數 μ_A 、 σ_A 、 μ_B 、 σ_B 及 P_A 、 P_B 可

以 $1-P_A$ 獲得。

假如我們知道圖 3-4 兩常態分佈的每一個資料所屬族群，計算上述五個參數將可依下列公式求得：

$$\mu = (\chi_1 + \chi_2 + \dots + \chi_n) / n$$

$$\sigma^2 = (\chi_1 - \mu)^2 + (\chi_2 - \mu)^2 + \dots + (\chi_n - \mu)^2 / n - 1$$

上式中， $\chi_1 + \chi_2 + \dots + \chi_n$ 為 A, B 分佈之樣本數，若要預估第五個參數 P_A ，只要取得叢集 A 的分佈即可。

假如已知上述五個參數之下，給予任何一個值 x 而其屬於叢集 A 的機率可以下列公式表示：

$$P[A | \chi] = P[\chi | A] \cdot P[A] / P[\chi] = f(x; \mu_A, \sigma_A) P_A / P[\chi]$$

$$f(x; \mu_A, \sigma_A) = \frac{1}{\sqrt{2\pi} \sigma_A} e^{-\frac{1}{2\sigma_A^2} \|x - \mu_A\|^2}$$

$f(x; \mu_A, \sigma_A)$ 為叢集 A 的常態分配函式。

• EM 演算法 (The EM Algorithm)

由上面之假設是已知資料集的資料分佈，以及五個參數的值，然而目前的問題是當以上兩種情況都不明的時候該如何做呢？因此採用了類似之前所提的 k-means 叢集演算法之重複尋找質心之方式 (Mitchell, 1997)。以下則分別說明 EM 演算法其執行步驟。

步驟一：

評估步驟(Estimation Step)：首先假設叢集參數之平均值 μ_i 、標準偏差 σ_i 即可推算出每點到此叢集 q^i 之機率 p_i ，且 $q^i = 1$ 。'i'代表假設第 i 個族群。因此利用上述的 μ_i 、 σ_i 及 p_i 之初始值來計算每個資料點(instance) χ 之叢集機率。利用所得到的機率，及貢獻度(Responsibility)再重複預測 μ_i 、 σ_i 及 p_i 之值。

所謂貢獻度之定義為：在一資料集中每一資料點 χ 在每個叢集 q^i 之隸屬程度。對任一點 χ 的貢獻度以下列式子表示：

$$\text{貢獻度 } h_n^i = \frac{\frac{\pi_i}{\sigma_i^d} e^{-\frac{1}{2\sigma_i^2} \|x_n - \mu_i\|^2}}{\sum_j \frac{\pi_j}{\sigma_j^d} e^{-\frac{1}{2\sigma_j^2} \|x_n - \mu_j\|^2}}$$

簡單的貝式公式如下：

$$h_n^i \equiv P(q_n^i = 1 | x_n) = \frac{P(q_n^i = 1)P(x_n | q_n^i = 1)}{\sum_j P(q_n^j = 1)P(x_n | q_n^j = 1)}$$

步驟二：

最大化(Maximization Step)：在此步驟依每一個資料點貢獻度的大小，重新計算叢集之平均值 μ 及標準偏差 σ ，並以新的平均值 μ 當叢集中心。

平均值則依每一資料點在該叢集的貢獻度而定其公式如下：

$$\mu_i = \frac{\sum_n h_n^i x_n}{\sum_n h_n^i}$$

接下來我們將考慮如何終止此重複的演算。在 k-means 的演算法中，將重複進行

尋找新質心直到群集參數不再變動為止，此時最終點為一固定點。而在 EM 的演算法中，並不是那麼簡單，我們是藉由資料集之 μ_i 、 σ_i 及 p_i 之值，計算出某個叢集 q_i 的概似值(Likelihood)，而其值可由各資料的機率相乘而得。其公式如下：

$$P(X_n | q_i = 1) = \prod_{k=1}^n P(X_k | q_i = 1)$$

上述的概似表示法並非介於 0 與 1 之間的機率值，而其值卻可表示整各叢集的好壞，而當 X_k 的每一點使得 $P(X_n | q_i=1)$ 最大時，此稱之為最大概似法(Maximum likelihood)。

一般為了方便計算，將以對數方式加以表示。本研究所使用之 Index Miner 其最後叢集結果，則採取對數之最概似法(Log Maximum likelihood)。其數學表示法如下：

$$\text{Log } P(X_n | q_i = 1) = \text{Log } \prod_{k=1}^n P(X_k | q_i = 1)$$

第四章 研究設備與方法

本研究之目的在於對門診就診病患之屬性，以資料探勘之叢集方式做一深入之探討。其重要屬性之取得，初期將以統計之線形圖，並依各曲線斜率之變化分析，找出與醫師看診時間可能有關之屬性。如病患本身就診之屬性：包括新病患、初診、複診病患、性別與年齡等，以及提供予病患服務之屬性：如醫師、科別、就診時間、時段與星期等。其次再將可能會影響醫師看診時間之屬性，以資料探勘之叢集分析方法，以建立各醫師與病患看診時間模式，以作為病患預約掛號看診時間預估之參考。

4.1 資料收集

因醫院醫療資訊系統資訊化程度不同，資料之收集方式將有所差別，本研究因某區域醫院其電腦化之程度較高，故其資料之取得乃以電腦資料為主，醫師看診時間之取得以一筆交易完成後取得電腦主機時間為準。因醫師為連續看診，故每一位病患之看診時間為：目前病患看診結束電腦時間，減去前一位病患看診結束電腦時間。對於第一位看診病患因醫師實際看診時間之差異，故第一位病患不予計算。為有效獲取正確之資料，加上資料探勘時亦需龐大的資料做為分析。本研究以主從式之架構，在 PC 端以 SQL 直接至資訊系統主機取得相關資病患就診屬性，資料範圍為整個月之門診就診資料，約五萬多筆，其欄位資料包括病歷號、新病患、初複診、就診序號、就診預定時間、實際醫師看診完成時間、性別、年齡、疾病別與疾病名稱。而目前各醫院之醫師排班均以週為主，故一個月可有四次之重複性資料取得，以確保資料之可用性。

資料收集期間：以 2001 年 10 月份之門診就診資料為主。

研究對象：為資料蒐集期間，至該院就診的門診掛號病患。

4.2 研究材料與設備

本研究所採用之資料來源，以北部某區域教學醫院一個月之門診電腦資料，為本研究範圍。所使用之軟體包括 Oracle Database 及 SQL Plus，此作為原始資料取得與初步過濾，其次用 Excel 再次過濾資料並產生所需欄位。然為便於研究與資料型態之轉變，將 Excel 的資料集轉至 Access 以檔案方式儲存資料，此兩含有相同資料內容之資料集將作為以後統計與叢集分析之資料來源。

針對與醫師看診有關的屬性以 SPSS 10.0 版，開啟 Excel 檔案並以線性圖斜率之關係，以了解整體統計之相關分析以獲取所需之屬性特性。資料探勘叢集分析則利用 Index Software Inc.之 Index Miner 軟體，透過 ODBC 與 Access 資料庫連接，以獲取各屬性與醫師看診有關之叢集模式。

4.3 研究方法與架構

本研究首先將以一般統計方法利用線形圖之斜率變化與面積大小，以了解二項式變數間之關係。此線形圖一方面可以有效的同時表示多個變數值，另外其斜率之變化可清楚的表示 X 軸與 Y 軸之趨勢變化之不同。例如圖 4-1 性別與醫師看診時間之關係圖，兩個曲線圖分別代表男與女，其 Y 軸為就診人數，若只差 Y 軸之位移(offset)兩曲線就重疊在一起，我們則可以視為男與女之醫師看診時間沒有差異。依此方式以探討與醫師看診時間有關之資料屬性，以作為叢集之資料充實以及其結果判定之參考依據。

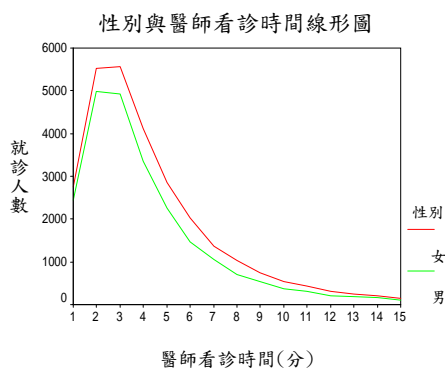


圖 4-1 性別與醫師看診時間之關係圖

例如各星期看診病患人數、個別醫師看診時間、科別與病患初複診之資料等。然

多屬性間之族群關係與相似性則以資料探勘之叢集方法，以建立醫師看診時間叢集模式，以作為預約掛號醫師看診時間之參考標準。然將醫師看診時間模式分析結果，串到排隊理論將是複雜的課題，加上預約與非預約人數之比例及爽約人數之不確定性，在本研究之時間內，恐難以達成。故本研究只以病患之屬性，探討醫師看診時間模式，後段之等候理論以預測看診時間將不予探討。本研究之研究流程如圖 4.2

圖 4.2 為病患就診，醫師看診時間預測模式：病患可自醫院所提供之掛號系統，採用現場或預約掛號，經病患所提供之資料及醫院醫療資訊系統(HIS)所提供相關的資料庫，以 SQL 獲取相關的就醫屬性，依需要篩選、編碼及轉換與醫師看診時間有關之資料，如醫師、科別、星期、時段等。根據上述屬性，依叢集理論之 EM (Estimation-Maximization)演算法，給予適當之參數與屬性後，最後依叢集結果之 Maximum Likelihood 與領域經驗之評估，歸納出醫師看診時間 (ΔT)之叢集結果。

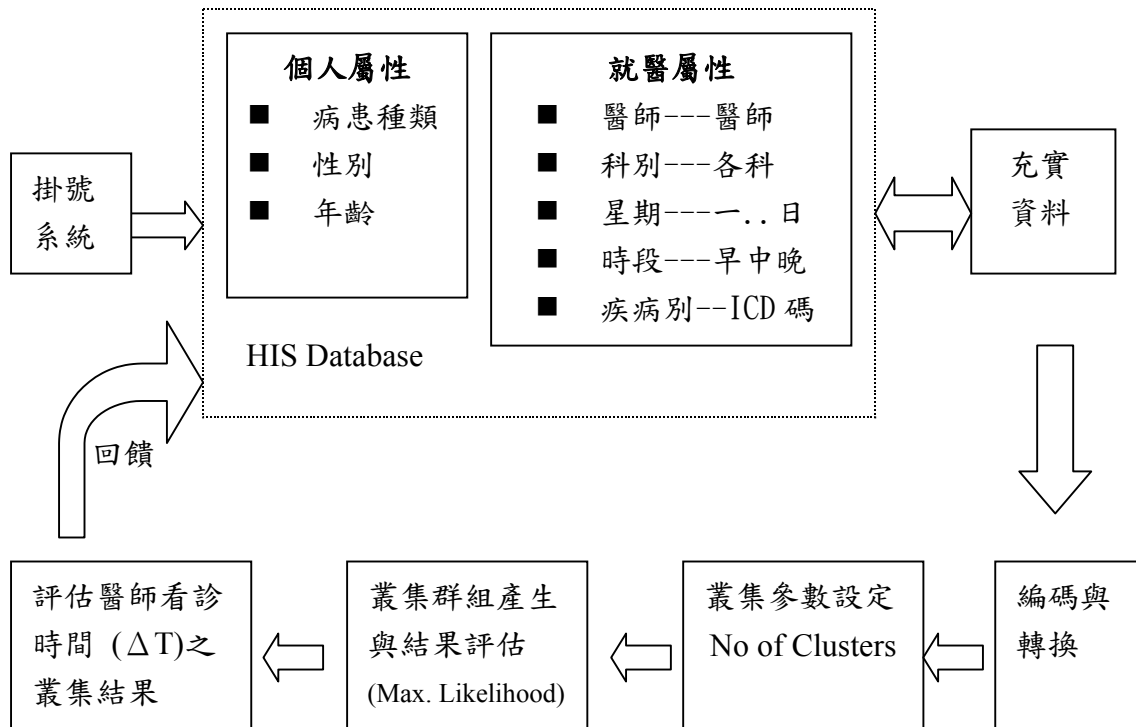


圖 4.2 以叢集方式評估醫師看診時間之架構

4.4 研究步驟與流程

圖 4.3 為叢集方式評估醫師看診時間研究流程，整個流程中主要有兩個判定點，其一為判斷資料是否充實？此一階段之重點在於完成資料的收集、過濾、篩選、及必要之編碼與轉換。另一為評估預測結果的正確性，主要依門診病患之相關屬性與叢集結果，以判定資料之正確性。

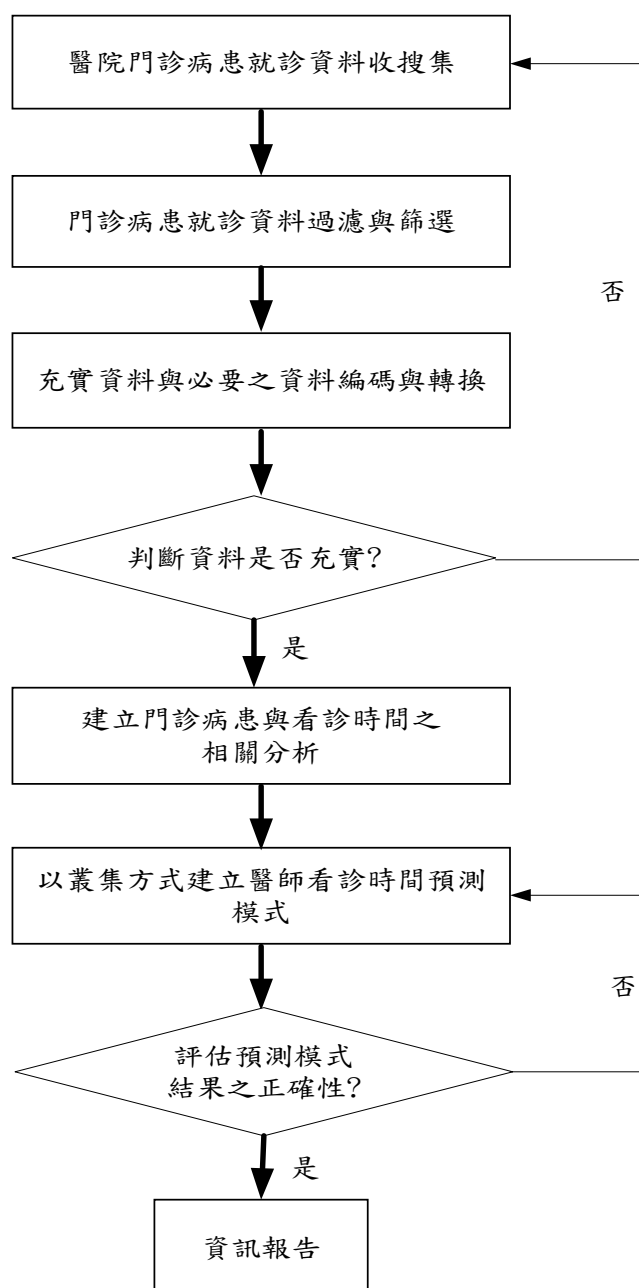


圖 4.3 以叢集方式評估醫師看診時間研究流程

4.5 Index Miner 之使用

本研究採用 Index Miner 作為資料探勘之工具，在此檢述其使用方式：

1. 啟動 Index Miner：首先須安裝 JDK1.3 以上之 Java 開發工具，之後才可啟動此軟體。
2. 資料連接：因 Index Miner 提供多種資料載入方法，本研究採用 Open Database 之方式，此方式須先設定好 ODBC 與資料庫之連接，之後在 Data Processing 中選取 Open Database 將其放置於 Project 之畫面中。
3. 資料載入：於 Project 中，將游標指在 Open Database 按右鍵，選取「Load data」，其參數設定如圖 4-4 所示。此時於「Query」中輸入所需 SQL 查詢語言。
4. Scheme 之選擇：在 Clustering 中選取 EM 將其放至於 Project 中。
5. 資料與方法之關聯：於 Project 中選擇 Open Database 使其產生一箭頭指向 EM。如圖 4-4 左邊之紅色箭頭。
6. 參數設定：在 EM 之小圖示下按右鍵，如圖 4-5 於此可設定執行時所需參數，一般只設定叢集數目即可。
7. 執行叢集運算：圖 4-6 選擇「Start」即可。



圖 4-4 Index Miner 資料載入方式

在圖 4-5 中，“-1”代表叢集數是由系統設定。“max. iterations”表示最大重複

執行次數，若未達 minStdDev 所設定值，則系統將以此設定值之大小，作為重複執行次數。若系統未達最大重複執行次數，但已達 minStdDev 所設定值，系統將停止運算。“seed”則為任意種子之數目。

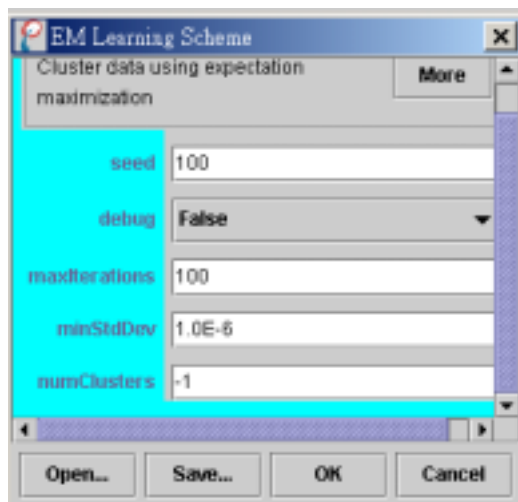


圖 4-5 參數設定

圖 4-6 為 EM 之屬性選擇與執行，若欲忽略不需要之屬性，可由「Ignore attributes」中選擇，如圖 4-6 之右邊所示。當參數與屬性選擇好後，按「start」即可開始從集之運算。

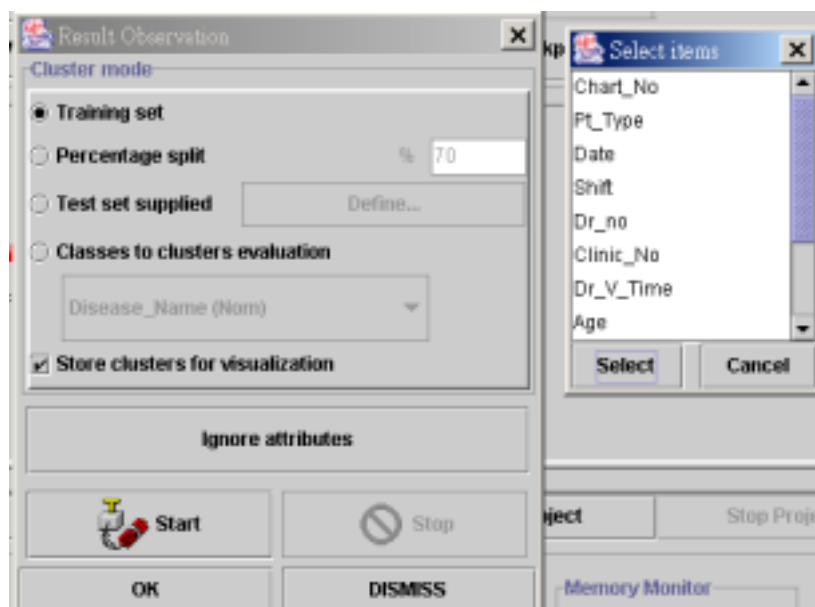


圖 4-6 EM 之參數設定與執行

第五章 研究結果

依據第三章之研究步驟與方法，本章將整個研究結果先以統計之方法，對於門診病患之樣本資料作整體性之分析。其次再建立病患就診屬性與醫師看診時間之關係。最後則以看診人數最多的內、外、婦、兒及神經內科醫師以叢集方式，建立個別醫師看診時間之叢集模式，並評估叢集模式之正確性，及其代表之意義。

5.1 門診病患資料描述與樣本資料分析

為了有效瞭解整個資料之特性，以作為叢集分析之前處理，下列為本研究資料之基本統計分析值。包括：各科看診人數，各醫師每月看診人數，該月之所有醫師看診時間分佈圖。

圖 5-1 為各科看診人數直條圖，其中人數最多者為一般內科，其次為小兒科，再次為婦產科及家醫科。其餘次專科部分包括骨科、皮膚科、眼科、神經內科、泌尿科、耳鼻喉科、美容科、腎臟科及直腸外科。

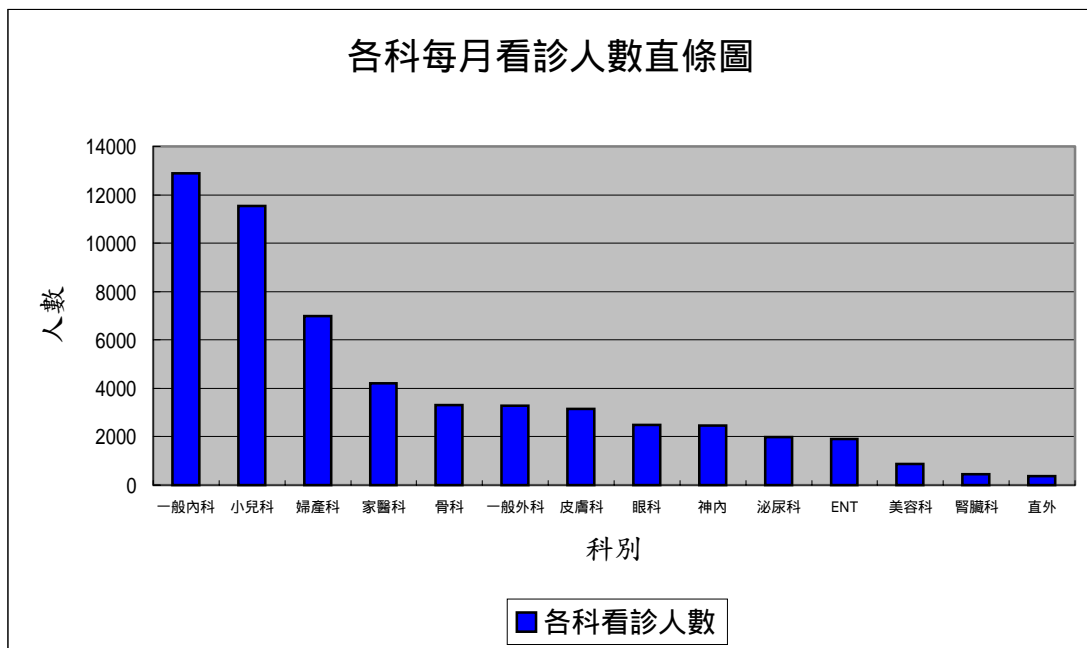


圖 5-1 為各科看診人數直條圖

圖 5-2 為各醫師每月看診人數之直條圖，為有效分析病患就診時醫師看診時間，此圖以該月醫師看診病患人次較多之醫師為主。而看診人數最多的醫師為 C28，B84，C40 以上三位均為小兒科醫師，而該科為該院第二大病患族群。

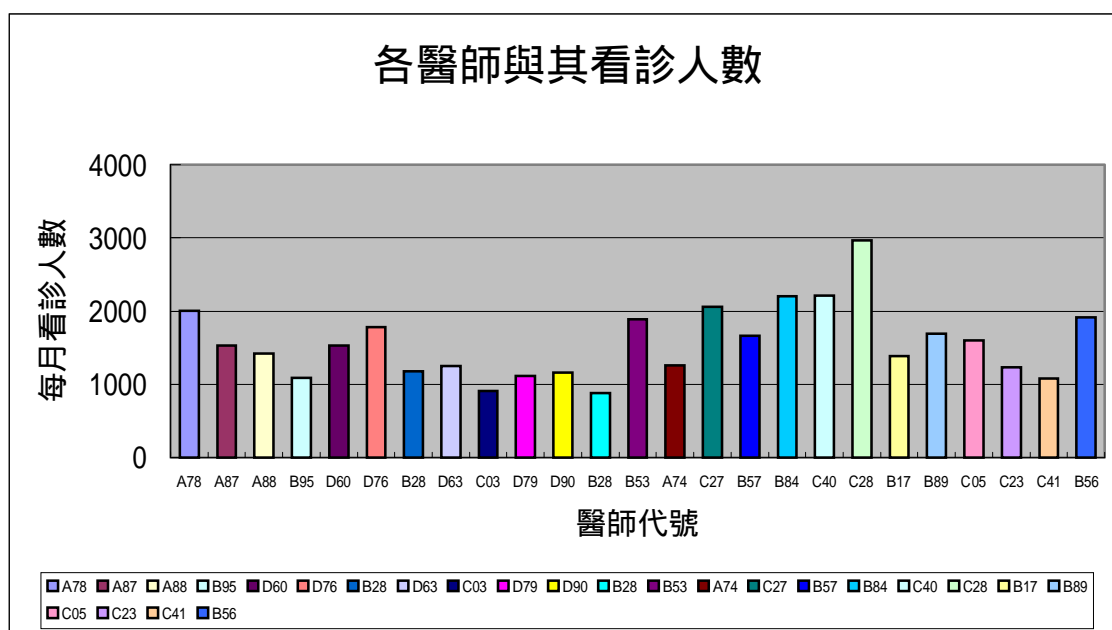


圖 5-2 為各醫師每月看診人數之直條圖

圖 5-3 為該月之所有醫師看診時間分佈圖，醫師平均看診時間為 4.5 分鐘，標準差為 3.34 分鐘。由圖可知大部份的門診病患，醫師的看診時間都集中在 2.5 至 5 分鐘之間。時間超過 20 分鐘以上之資料將與刪除，其原因包括醫師會診，或其他非看診因素導致該病患看診時間有中斷，使整個時間超過 20 分鐘以上。

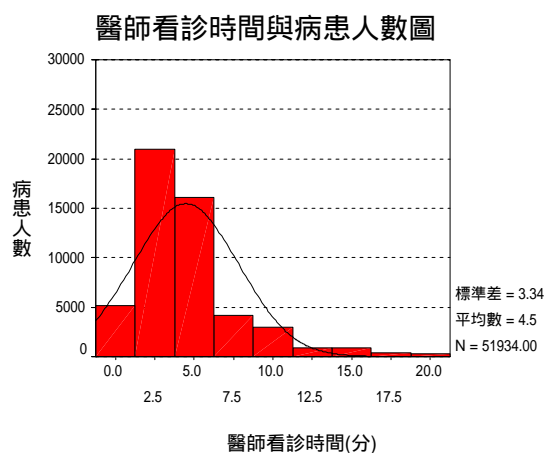
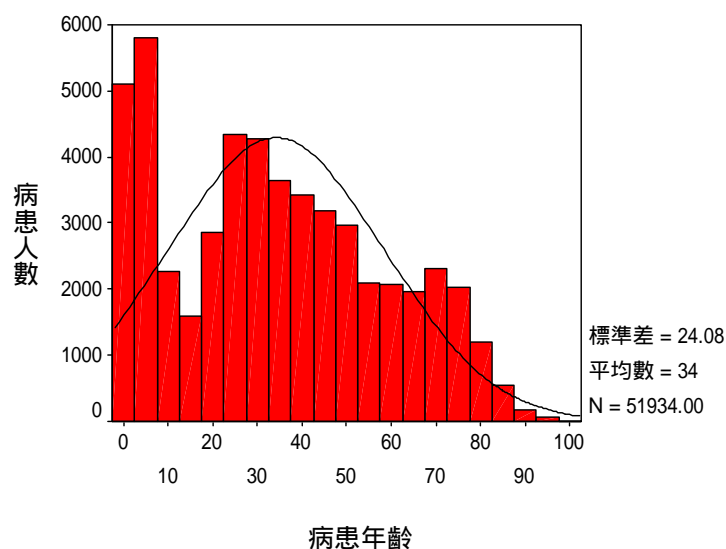


圖 5-3 為該月之所有醫師看診時間分佈圖

圖 5-4 為病患年齡與就診人數分佈圖，資料為一個月之門診病患人數，主要有三大族群分佈，其一為 0~5 歲之病患，另一族群為 25~50 歲之病患。最後一族群為 70 歲之病患。由此分佈圖來看，病患屬於幼兒及年輕之族群居多。



病患年齡與就診人數分佈圖

圖 5-4 為病患年齡與就診人數分佈圖

根據上述分析結果可知，一般內科、小兒科及婦產科此三大科別佔就診人數約五分之三。其次以醫師看診人數來看，每位醫師差異很大，其中又以小兒科醫師之看診病患人數最多，次專科則以神經內科病患人數最多。

5.2 門診病患就診屬性與醫師看診時間之關係

然而是否醫師看診時間與病患人數、科別有關呢？以下則探討病患就診屬性與醫師看診時間的關係，包括：星期、看診時段(早、中、晚)、病患種類(新病患、初診、複診)及各科別醫師等。另外依需要作必要之資料充實、編碼與轉換，再以統計之計數方法，對病患就診屬性與醫師看診時間之關係，做一線形圖之分析。於此，發現醫師看診時間超過 15 分鐘至 20 分鐘以上並無差異，為使線形圖中之相關曲線能表現清楚，超過 15 分鐘至 20 分鐘以上之資料將與省略。

• 醫師看診時間與時段之關係：

早、午、晚三個時段之病患人數百分比分別為 39：24：37 如表 5-1，看診時間範圍為一分鐘至二十分鐘，超過二十分以上者，因醫師會診或其他因素不予列入計算。大部分醫師對於病患的看診時間介於一分鐘至六分鐘間。時段對醫師看診時間而言，早、晚二個時段並無差異。午診與早、晚診則有所不同。

表 5-1 病患看診時間與看診時段的關係

		看診時段		
		早	中	晚
		病患人數	病患人數	病患人數
1	分鐘	2140	1209	1850
2		4045	2464	3990
3		4009	2443	4043
4		2928	1718	2827
5		1973	1238	1902
6		1356	892	1257
7		876	657	886
8		664	441	636
9		486	336	466
10		357	246	312
11		251	208	283
12		210	146	150
13		162	117	153
14		143	108	122
15		96	89	81
16		86	77	74
17		99	61	71
18		68	64	45
19		57	67	42
20		55	51	51
總	個數	20061	12632	19241

圖5-5為醫師看診時間與時段的關係，由此圖中之曲線可發現早、晚幾乎相同，而午診看診人數較少。而共通之現象為醫師看診時間，以二至三分鐘之人數最多，看診時間越久病患人數相對減少。

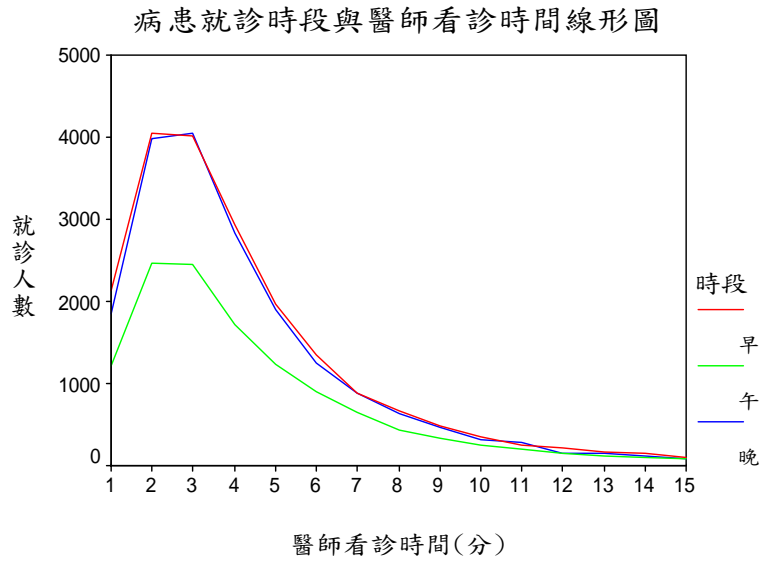


圖 5-5 為醫師看診時間與時段的關係

• **醫師看診時間與星期之關係**

圖5-6為病患就診星期與醫師看診時間的關係，此圖表可看出在星期一、二兩天病患人數最多，而星期日則明顯的減少。醫師看診時間，以二至三分鐘之人數最多。星期日則以二至四分鐘之病患人數較多。

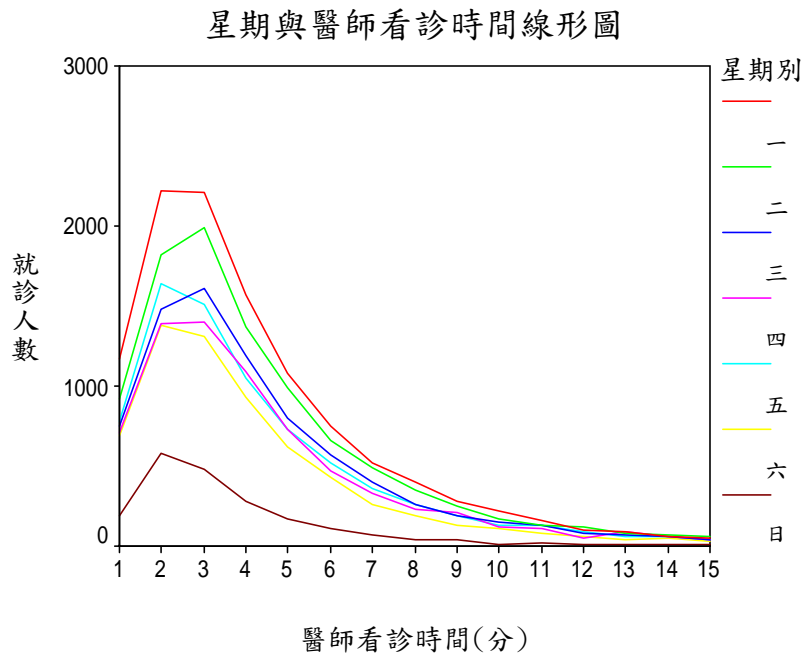


圖 5-6 病患就診星期與醫師看診時間的關係

每週一至週日病患人數如表 5-2，其中以每週一病患人數最多，週六下午及週日因縮短開診故病患總人數較少，尤以週日人數最少。然而因個別醫師排班之不同，其看診人數與全院之病患看診人數並無關係。如 C28 醫師於週日看診病患為該週最多。’—部分表示該醫師於該星期未排班看診。有底線部份表示該醫師最多病患之星期。

表 5-2 門診病患人數與星期別之關係

	門診病患	A78 醫師	A86 醫師	C28 醫師	D63 醫師	B84 醫師
星期	總人數	病患人數	病患人數	病患人數	病患人數	病患人數
一	<u>11071</u>	346	264	530	209	<u>597</u>
二	9654	397	<u>587</u>	564	<u>241</u>	432
三	7922	288	152	310	158	178
四	7164	<u>423</u>	313	477	196	315
五	7733	--	216	242	165	329
六	6354	240	30	80	167	250
日	2036	--	38	<u>619</u>	43	--
總計數	51934	1694	1600	2822	1179	2101

註： "--" 表示醫師於該星期未排班看診

以上資料為一個月之統計量

- **病患種類與醫師看診時間之關係**

圖5-7為病患種類與醫師看診時間的關係。新病患：為該病患首次到該院來就診，初診：為該病患已有病歷資料，但首次由此醫師看診，複診：為病患已有病歷資料且由同一位醫師於同一科別看診兩次以上。此圖可知，主要病患族群為初診病患再次為複診病患，此兩族群之病患看診時間主要集中在二至三分鐘。而新病患所佔人數最少，看診人數亦集中於二至六分鐘之時段。

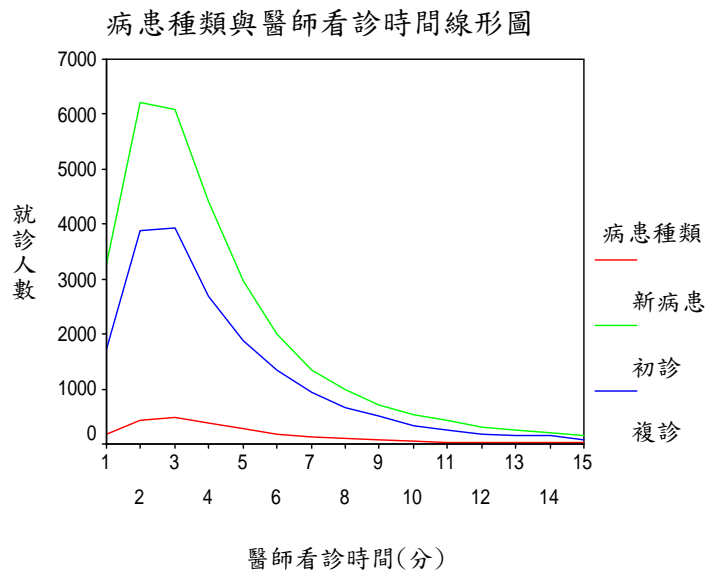


圖 5-7 為病患種類與醫師看診時間的關係

- **科別與醫師看診時間之關係**

圖 5-8 為醫院之主要科別內、外、婦及小兒科四科為主之線形圖分析，並以門診人數較多的醫師為主。由此圖可以看出，不同科別其醫師看診時間將有所不同，尤其以婦產科與其他科別差異最大，另外由圖可看出，不論哪一科，就診人數最多的醫師其看診時間亦相對減少。以內科及外科而言，看診人數與看診時間呈相反之關係，看診人數越多，醫師看診時間越短，但也非全然如此。婦產科與小兒科則無相對的明顯關係，主要與個別醫師之不同，看診時間差異極大。

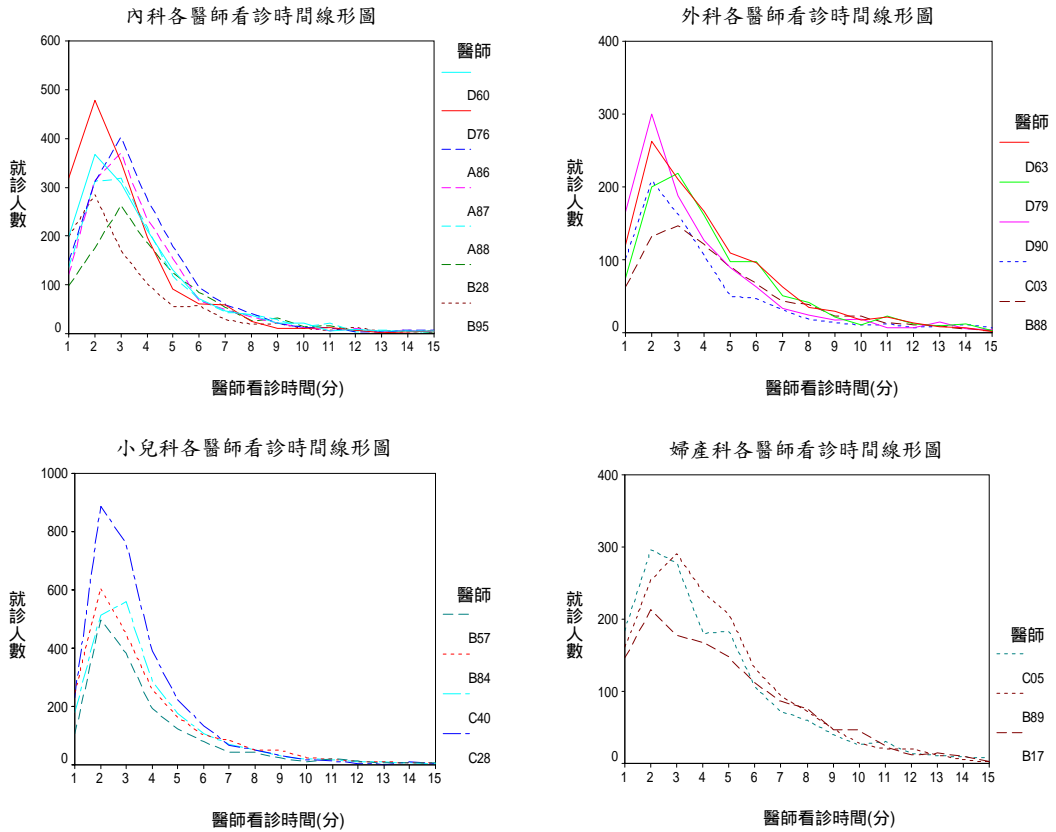


圖5-8為內、外、婦、兒四科與醫師看診時間的關係

• 病患性別、年齡與醫師看診時間之關係

圖5-9為病患性別及年齡與醫師看診時間的關係。由左圖可知性別之曲線差異不大。而年齡則因各年齡層之不同差異極大，就診人數最多的為綠色，醫師看診時間為二分鐘，其次為藍色，醫師看診時間為三分鐘，再次為粉紅色，醫師看診時間為四分鐘。年齡層集中於5，30及70歲左右。

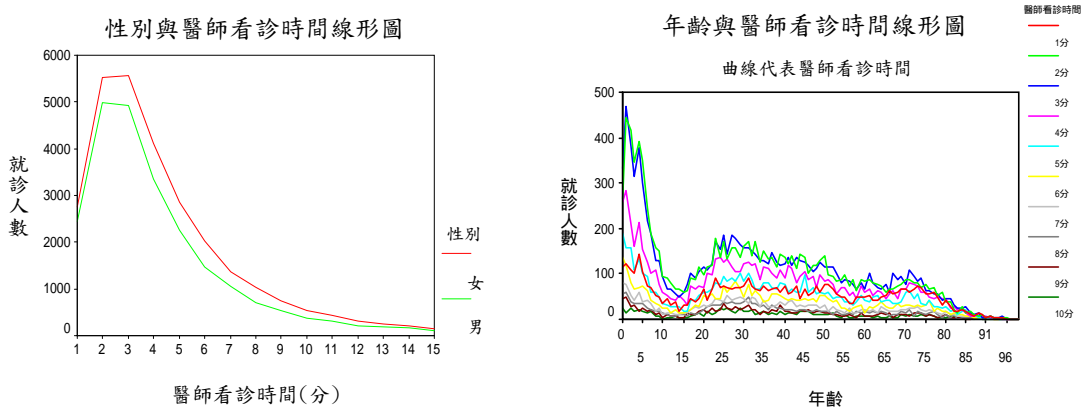


圖5-9病患性別、年齡與醫師看診時間之關係

醫師看診時間與病患各屬性之關係歸納如表5-3

由表 5-3 之屬性可知，整體而言：時段別、星期別、病患種類、科別與年齡均會影響醫師看診之時間，而性別上並無太大差異。因各科之特性不同其看診時間亦有所不同，而對同一科別之醫師看診時間方面亦有不同，因醫師之個別差異為影響醫師看診時間之主要因素。故在以下之資料探勘分析部分，將以個別醫師為分析對象。而實際上因門診就診之特性使病患掛號後為單一醫師服務之特性，該看診醫師與病患之互動關係，將影響醫師看診時間與掛此醫師之病患等候時間。

表 5-3 病患各屬性與醫師看診時間之關係

屬性	醫師看診時間
時段別	午診與早、晚診有顯著差異，早診與晚診則無明顯差距，因為午診的門診量較其他時段低，是否代表隨著病患數量的減少而醫師調慢看診速度，或醫師等候病患時間變長。
星期別	表示星期會影響醫師看診時間，尤其以週日最為明顯，相對於病患人數，週日亦為最少。然對個別醫師言，卻不一定。
病患種類	新病患所佔人數最少，看診時間與初複診有極大之差異。因新病患病歷記載之基本資料必須一一詢問。
科別	同科別不同醫師看診時間將有所不同，尤其以婦產科其與其他科別差異最大。不同科別醫師看診時間亦有所不同。
性別	男女性別整體性而言並無差異。
年齡	年齡則因各年齡層之不同差異極大。

5.3 以資料探勘之 EM 法分析病患屬性與看診時間之關係

由上述資料之初步統計分析，個別醫師將是影響看診時間之主要因素，以下將會影響醫師看診時間之屬性，包括：時段別、星期別、病患種類、科別與年齡等列入分析之資料，性別於上述二項式分析中並無差異，然而為了解女性族群之看病特性，在此亦加入叢集之分析，至於疾病碼 (ICD-9 Code)與看診時間有密切的關係故亦加入分析。另外對於病歷號碼與疾病名稱僅為參考用並不列入叢集分析資料內。看診人數亦為重要屬性，然因其數量即為每一筆資料的總合，亦即為當月該醫師看診總門診人數，以含於資料內。依上述屬性再分別以資料探勘之分析方法，以了解多維屬性之間與醫師看診時間之相互關係。並以叢集之 EM 法，以了解各屬性間與醫師看診時間之模式。

5.4 叢集法之參數設定

以下分析將以門診人數較多之內、外、婦、兒及次專科等個別醫師作為分析之對象，因其初複診之人數較多，隔週就診之比例較高，對複診資料之取得較為容易，另外因生病於三至七天內再次就診之比率亦最高，此研究取一個月之資料為分析對象其正確性亦較高。初次選擇的欄位屬性如下：病歷號碼(Char-no)，看診日期(Date)，病患種類(PT Type):新、初、複診病患，時段別(Shift)，醫師看診時間(Dr. Service Time)，年齡(Age)，性別(Sex)，疾病碼(ICD-9 Code)。欄位屬性中除醫師看診時間及年齡為數字型態外，其餘均為文字型態。忽略之屬性為疾病名稱(Disease Name)，其主要目的為依疾病碼之結果判定其疾病為何，故不放入分析資料內。看診時間以四捨五入計，未滿一分鐘者以一分鐘計。

以下將個別醫師相關資料分述如下，其科別與醫師包含:神經內科 A78 醫師，一般內科 A86 醫師，小兒科 C28 醫師，一般外科 D63 醫師及婦產科 B89 醫師。

表 5-4 個別醫師新病患與初、複診病患資料

新病患與初複診病患資料表					
新病患與 初複診病患	人數				
	神經內科 A78 醫師	一般內科 A86 醫師	小兒科 C28 醫師	一般外科 D63 醫師	婦產科 B89 醫師
新病患	22	73	39	57	62
初診病患	837	1191	1114	535	835
複診病患	835	326	1669	587	703

表 5-5 看診時段與人數資料

看診時段，人數與不同科別醫師資料					
看診時段 (早，晚)	人數				
	神經內科 A78 醫師	一般內科 A86 醫師	小兒科 C28 醫師	一般外科 D63 醫師	婦產科 B89 醫師
早	1127	900	305	383	729
午	無看診	197	901	359	153
晚	567	493	1616	437	718

以下將以叢集法來分析病患就診屬性與個別醫師看診時間之關係，其參數設定如下：

叢集之 Scheme：此採用 EM 法。

叢集個數：首次先取-1，表示由軟體系統自行決定最佳結果。爾後則參考首次結果及資料特性，由使用者自行決定所需叢集個數。叢集分析結果則依實際資料之群族特性，領域資訊與最大可能性(Maximum Likelihood)之值決定。相關屬性之遞減亦可依叢集之結果，由使用者自行決定，以得到更有意義的分析資訊。

最大重複次數(Max. Iterations)：100 次

種子(Seed)：100

最小標準偏差(Min. Std. Dev)：1.0 E-6

圖 5-10 為 Index Miner 之實際操作之畫面，右邊之小視窗則為參數設畫面。

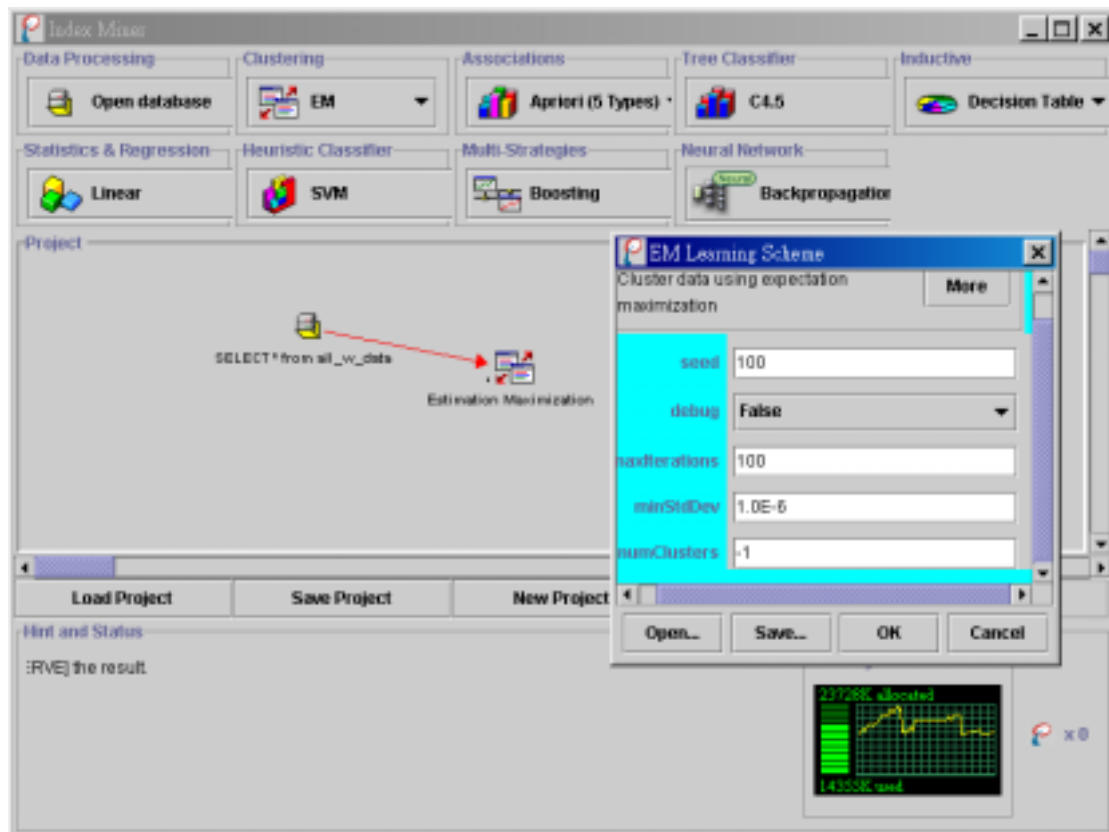


圖5-10為Index Mining 之參數設定值

5.5 叢集分析之結果

依據上述之結果，資料探勘叢集分析部分，將以個別醫師為研究參考對象，共有五位。其最後之叢集分析將參考絕對值最大之 Log Likelihood，並與之前分析之病患屬性與醫師看診時間結果，以作為獲取最有代表意義之叢集事實現象與看診時間關係之結果。以下則為個別醫師 EM 分析結果，其屬性包括：病患種類、星期別、時段別、年齡、性別與主診斷等。

• 一般內科 A86 醫師 EM 叢集結果

一般內科 A86 醫師的主要病患集中在叢集 3，佔總比例的 64%。此一叢集主要病患之疾病別為本態性高血壓(Essential Hypertension)，年齡約 63 歲，主要看診時段集中在早上，複診病患居多。叢集 2 出現於週日，此醫師週六、日看診人數較少，週二病患人數最多，但看診時間並未因看診人數之增加而相對減少，平均看診時間為 3.67 分。另外新病患主要集中於週日下午與晚上看診，此可能與週日一般診所並未開診有關，尤其晚診人數最多，主要疾病別為急性呼吸道感染。另外叢集 1 佔 13%，主要疾病別為本態性高血壓(Essential Hypertension)，其平均年齡為 55 歲，醫師看診時間平均為 6.45 分，相關研究報告指出，高血壓與年齡關係，大多出現在 40 歲以後，我國 40 歲以上的人口中，高血壓的盛行率高達 18%，55 歲達 26% (台灣醫學會 1992)。由此兩族群之本態性高血壓病患來看，年紀較輕者，主要是初發病患及病齡較短，且新病患與初診病患比率較高，病患與醫師間之互動性較高，而長期之病患因病情已控制，固定回診檢查，醫師與病患間已有固定之互動模式，所需時間較短且固定。

表 5-6 一般內科 A86 醫師之 EM 叢集結果

叢集數目 = 4		Log likelihood: -9.2			
屬性	單位	叢集 0	叢集 1	叢集 2	叢集 3
比率	%	5	13	18	64
病患種類					
(新病患:初診:複診)	數量	2:3:0	11:163:30	49:229:11	10:283:726
星期:人數/總人數	數量	6:43/74	5:153/204	7:102/290	2:238/1020
時段(早、中、晚)	數量	72:0:0	19:165:20	117:31:143	690:0:330
醫師看診時間	分	2.88	6.45	3.25	3.67
醫師看診時間	標準差	1.20	1.50	1.47	1.92
年齡	平均值	61	55	29	63
年齡	標準差	18	18	11	13
性別(男:女)	數量	26:47	103:101	153:136	512:508
主診斷與個數	診斷碼:數量	4349:6	4019:77	4659:105	4019:501

• 神經內科 A78 醫師 EM 叢集結果

神經內科 A78 醫師，經叢集以 EM 之方式分析門診病患資料，因為屬於專科醫師，病人的主要疾病診斷範圍較一般內科集中，複診且重複就診之病患亦較多，故其看診時間之差異性較低，以四個叢集結果來看，平均看診時間從 2.99 分到 3.55 分，而初複診病患醫師花費看診時間並無太大差異。以年齡層來看，平均 40 與 42 歲之病患族群看診時間較 69 及 71 歲之病患族群，看診時間較長。另一方面來看，主要診斷為 4349 (腦動脈阻塞)其年齡層皆 70 歲左右之族群佔 59%，男性多餘女性且集中在早上時間看診。另外兩族群，主診斷為 7840 (頭痛)以平均 40 歲左右之病患族群為主，且女性多於男性。

表 5-7 神經內科 A78 醫師之 EM 叢集結果

叢集數目 = 4		Log likelihood: -7.37			
屬性	單位	叢集 0	叢集 1	叢集 2	叢集 3
比率	%	42	17	18	23
病 患 種 類					
(新病患:初診:複診)數量		1:351:346	1:122:136	5:140:152	1:201:164
星期:人數/總人數	數量	2:245/698	1:153/260	1:189/298	2:136/375
時段(早、晚)	數量	697:1	1:259	1:297	375:1
醫師看診時間	分	2.99	3.16	3.36	3.55
醫師看診時間	標準差	1.51	1.75	1.70	1.70
年齡	平均值	69	71	42	40
年齡	標準差	9	7	13	11
性別(男:女)	數量	380:318	149:111	122:176	155:220
主診斷與個數	診斷碼:數量	4349:170	4349:89	7840:32	7840:57

• 小兒科 C28 醫師 EM 叢集結果

小兒科 C28 醫師的主要病患集中在叢集 2 與 3，佔總比例的 66%。此一叢集主要病患之疾病別為過敏性鼻炎(Allergic Rhinitis Cause)，年齡約 4 歲，主要看診時段集中在晚上，且複診病患居多，分別佔 47%及 70%。因該區主要為工業區，大部份兒童之父母均為上班族，下班後才帶小孩看診之情況居多。醫師看診時間最短分別為 2.56 分及 2.62 分鐘，分別出現於週二及週日，該兩天病患人數較其他天多一倍，此醫師之看診時間會因看診人數之增加而相對減少。另外新病患於週日稍有增加但並不明顯。

表 5-8 小兒科 C28 醫師之 EM 叢集結果

叢集數目 = 5						
屬性	單位	叢集 0	叢集 1	叢集 2	叢集 3	叢集 4
比率	%	17	12	33	30	8
病患種類 (新病患:初診:複診)	數量	3:175:284	11:167:163	3:292:615	14:333:479	4:120:115
星期:人數/總人數	數量	1:113/463	1:90/341	2:230/910	7:356/826	2:170/239
時段(早、中、晚)	數量	2:98:363	64:163:114	1:1:909	0:626:199	236:0:3
醫師看診時間	分	5.54	4.10	2.56	2.62	3.66
醫師看診時間	標準差	2.80	2.21	0.92	0.02	1.90
年齡	平均值	5.3	0.0	3.6	4.1	3.7
年齡	標準差	3.2	0.0	2.0	2.5	2.6
性別(男:女)	數量	284:179	185:56	496:414	469:356	117:122
主診斷與個數	診斷碼:數量	4779:351	4779:145	4779:676	4779:595	4779:175

Log likelihood: -6.23 ; ICD-9 碼主診斷 4779 為:過敏性鼻炎(Allergic Rhinitis Cause)

• 一般外科 D63 醫師 EM 叢集結果

一般外科 D63 醫師的主要病患集中在叢集 0, 1, 3 佔總比例的 90%。此一叢集主要病患之疾病別為 Unspecified Breast Disorder, 年齡層介於 30 至 36 歲間, 以母體資料而言年齡層介於 30 至 42 歲間, 主要看診時段平分於早、中、晚三階段。初、複診病患比率相差不多。男性比女性多 15%。另外, 叢集 2 與 4 佔 10%, 主要疾病別為蜂窩組織及膿瘡(Other Cellulitis And Abscess), 其平均年齡為 35 歲及 42 歲, 男性比女性多 31%, 醫師看診時間平均為 11.6 分, 此為病患換藥之故。

表 5-9 一般外科 D63 醫師之 EM 叢集結果(其原始之叢集資料, 請參考附件

一)

叢集數目 = 5

屬性	單位	叢集 0	叢集 1	叢集 2	叢集 3	叢集 4
比率	%	34	29	4	27	6
病患種類						
(新病患:初診:複診)	數量	12:178:207	25:154:163	2:23:22	16:155:147	2:24:48
星期:人數/總人數	數量	2:222/398	6:152/341	5:28/46	1:191/317	4:21/74
時段(早、中、晚)	數量	0:0:398	341:0:0	5:41:0	0:317:0	35:0:38
醫師看診時間	分	3.35	3.62	11.87	3.82	11.47
醫師看診時間	標準差	1.79	1.87	4.01	2.11	3.22
年齡	平均值	30.10	36.06	35.44	33.02	41.66
年齡	標準差	17.47	20.10	17.85	20.48	21.19
性別(男:女)	數量	211:186	160:182	30:16	168:153	49:25
主診斷與個數	診斷碼:數量	6119:40	6119:27	6829:6	6119:28	6829:6

Log likelihood : -8.44

• 婦產科 B89 醫師 EM 叢集結果

婦產科 B89 醫師的主要病患集中在叢集 1 與 4，佔總比例的 60%。此一叢集主要病患之疾病別為產前檢查(V221)，年齡範圍在 28 至 31 歲間，主要看診時段集中在早上及晚上，複診病患居多，因為固定之產檢醫師與病患間，已有固定之互動模式，所需時間較短且固定，平均看診時間為 3.8 分。叢集 0 主要病患之疾病別更年期(6272)，平均年齡為 54 歲，看診時段集中於早上，平均看診時間為 3.5 分。叢集 2 主要病患之疾病別為子宮頸發炎(61610)，平均年齡為 33.4 歲，看診時段集中於晚上，複診病患居多，平均看診時間為 5.58 分。叢集 3 主要病患之疾病別為腹痛(7890)，平均年齡為 36.9 歲，複診病患居多，看診時段集中於早上，平均看診時間為 9.86 分。看診時間與看診人數沒有一定之關係，醫師與病患間，已有固定之互動模式，所需時間較短且固定。

表 5-10 婦產科 B89 醫師之 EM 叢集結果

叢集數目 = 5		Log likelihood: -7.2				
屬性	單位	叢集 0	叢集 1	叢集 2	叢集 3	叢集 4
比率	%	16	24	17	8	36
病患種類 (新病患:初診:複診)	數量	3:203:48	6:138:234	15:144:108	5:62:61	33:288:251
星期:人數/總人數	數量	1:152/255	2:203/378	3:104/266	1:51/128	4:230/571
時段(早、中、晚)	數量	152:14:88	16:0:362	0:0:266	114:13:0	445:125:0
醫師看診時間	分	3.50	3.88	5.58	9.86	3.78
醫師看診時間	標準差	1.80	1.93	3.88	3.31	1.94
年齡	平均值	54.2	28.2	33.4	36.9	31.2
年齡	標準差	8.4	3.7	10.0	12.1	8.1

性別(男:女)	數量	0:255	8:369	0:266	0:128	0:571
主診斷與個數	診斷碼:數量	6272:170	V221:173	61610:75	7890:20	V221:131

第六章 結論與討論

6.1 研究結論與討論

一般在醫院之等候流程中，從掛號、看診、批價到領藥，影響最大且最難控制的就是等候醫師看診。以人力資源及排隊理論之運用來看，掛號、批價與領藥此三個等候站，均為多人可同時服務，且服務性質相似之服務站，病患在此三站中之任何一站，僅需由其中任何一位服務人員服務即可，服務人員之多寡可依病患到院之門診量而適當的調整。另外，工作內容之變化性較少，服務人員動作模式與行為較為固定，因此醫院方面也可斟酌人力成本與病患可忍受之抱怨程度，依「人」的工作效率加以正確的計算出最有效的人力運用與安排。在另一方面來看，此三站亦可經由自動化與資訊化的改善，提供單位人員更大的生產力。然在醫師看診方面因病患疾病之不確定性與多樣性，加上醫師之個人經驗與特性似乎沒有辦法以一個固定常數之模式加以預測，而在目前醫療資訊系統之使用上，大部分均無討論到醫師之個別特性，與病患之習慣而產生之看診時間關係，而此正是本研究重點所在。故本研究採用資料探勘之叢集模式，恰好可以解決面對病患，與醫院提供醫療服務資源之多屬性問題的困擾。

因叢集之特性，使每一位醫師之病患，均依其屬性之相似性而得到多個叢集，對每一個叢集都有該醫師看診時間，與病患在該叢集下所佔之百分比。此醫師看診時間與比例將作為設定病患掛號時，依病患就診屬性而決定其就診時間之依據。對一位有穩定病患以醫師而言，因其病患之變動性較少，運用叢集分析時，其資料之取得間隔可訂較長，例如每半年或一年，其主要之參考變動指標，可依病患之等候時間加以決定，是否重新產生新的叢集資料，以便修正醫師之看診時間。若對一位新醫師，此方式須待一定資料量後，才有較正確之結果產生。

在總額預算制度下，為顧及醫療品質，在醫師看診方面，限定病患掛號人數，

將與醫師看診時間息息相關。若經由醫師看診資料，透過叢集分析結果，將可提供醫院對每一位醫師合理門診量之運作參考。

在叢集特性所發現之事實方面，依本研究之內科、外科及小兒科，各醫師看診時間統計之線形圖結果來看，依看診人數之多寡，可得到一個趨勢：看診人數越多，則醫師看診時間相對減少。然以一般內科與神經內科看診人數較多之醫師而言，若以叢集之各族群來看，其含有其他意義。此兩科之最大族群比例各佔 64%及 42%，病患年齡介於 63 至 69 歲間，且複診比率高。此現象經與醫師討論後認為：以一般慢性病患而言，因病患病情已控制，且為固定回診檢查，醫師與病患間已有固定之互動模式，所需時間較短且固定。所以看診時間短，是因為病患歸屬於此類族群，且病患人數亦較多之故。而另外之族群則與疾病好發年齡有關，如 40 至 55 歲台灣之高血壓盛行率達 26%；42 歲左右則有頭痛之病患族群出現，且女多於男。以上情況因為初次發病，故初期與醫師間之互動性較高，詢問之問題較多，所以醫師花較多診察時間。可見病患之屬性與醫師之特性，隱藏著很多未知的事實且有用的知識。以下依據上一章之研究結果有如下的發現：

- 雖然是同一科別，但醫師看診時間亦大不相同。
- 男女生之看診時間沒有明顯差別。
- 專科醫師其看診時間較為固定。
- 年紀在 30~40 歲間其看診時間均較長，經發現因一般民眾白天上班，故下班後才去醫院看夜診，所以在工業形態的縣市，其夜診病患以 40 歲及小兒科人數較多，而此年紀之病患及父母與醫師之互動詢問較多，故感覺上夜診之看診時間較長。
- 一般老人家均習慣於早上看診。

6.2 未來研究方向

根據本研究之結果，已完成整個看診流程之前半段，圖 6-1 之模式則作為未來推展至整各看診時間預測之流程，其最後結果將以病患就診之屬性，

加上等候理論以決定病患之就診序號及預估看診時間。

病患就診預約時間預測模式

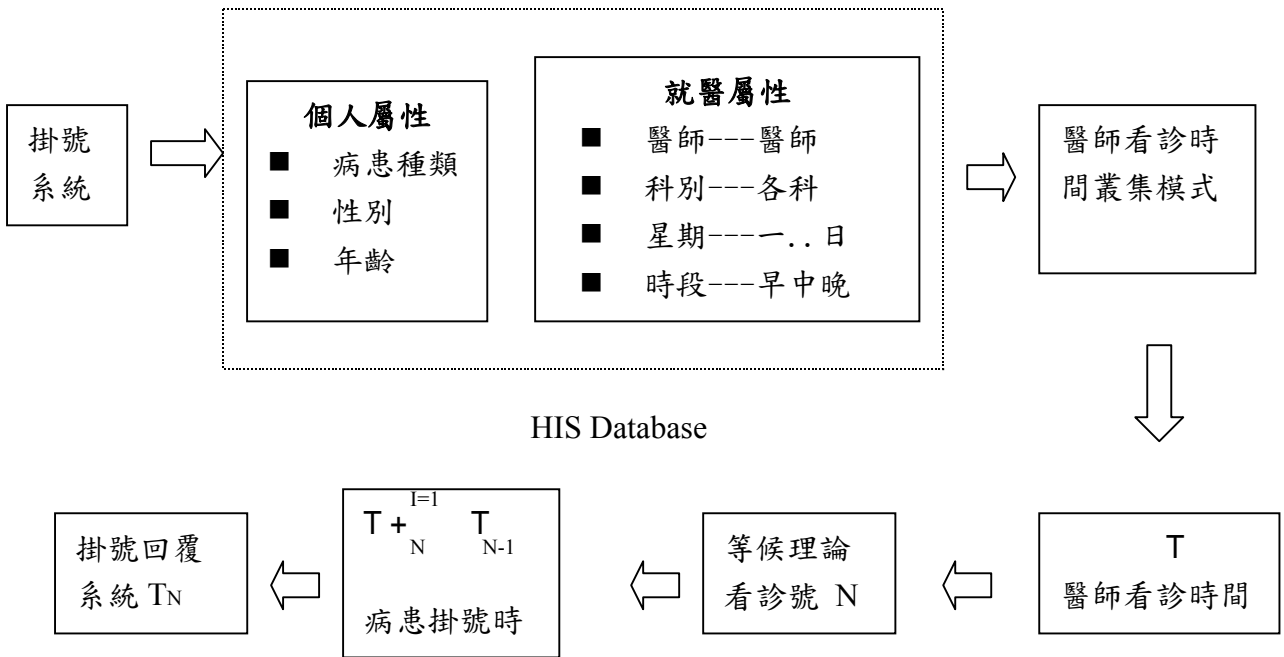


圖6-1 病患就診預約時間預測模式

圖 6-1 病患就診預約時間預測模式：病患可自醫院所提供之掛號系統，採現場或預約掛號，病患提供其基本資料後，醫院之醫療資訊系統(HIS)將提供相關就醫屬性，如醫師、科別、星期、時段等以便病患就醫。根據上述屬性，依叢集理論歸納出醫師看診時間 (ΔT)之叢集，再根據等候理論依病患以現場或預約掛號之方式及其事前之機率，決定病患之看診號(N)。T 代表各時段(早、中、晚)之開始看診時間，最後醫院之資訊系則以 T_N 回覆病患掛號預定就診時間。由上圖可知，醫師看診時間的模式預測將可作為病患等候時間之前導計劃，一般等候系統的結構。依等候之流程，顧客依序到達等候系統，進入系統後即加入等候線，在此根據等候管制規則(queue discipline)，先後接受服務設施(service facility)所提供的服務，服務完畢後即離開等候系統(作業研究，2001)。以一個已穩定之醫院而言，病患來源的數量可以之前之資料，如去年同期之病患就診人數，獲得相當之預測值。

參考資料

中文文獻

- 中華民國台灣醫學會，高血壓研究小組，「本土醫學資料庫之建立及衛生政策上之應用」，行政院衛生署八十一年度委託研究計畫研究報告，1992。
- 古政元，「門診等候時間縮短之研究—運用行動通訊技術」，國立中正大學資訊管理學系，2001。
- 行政院國軍退除役官兵輔導委員會-統計出版品，「榮民總醫院就醫病患狀況調查報告」，資料時間：88年 出版時間：89.12。
- 吳永順，「門診藥局等候時間最佳化之研究-----以成大醫院為例」，國立成功大學工業管理學系，1998。
- 吳佩環，「全民健保實施下影響門診病患選擇就醫層級之因素探討」，台北大學財政學系碩士班碩士論文，2000。
- 吳國禎，「資料探索在醫學資料庫之應用」，私立中原大學醫學工程學系碩士論文，2000。
- 侯幸雨，「應用模擬技術探討台灣醫院門診預約掛號系統」，國立中正大學企業管理研究所，1998。
- 胡國岱，「醫院門診掛號時間之研究」，國立中正大學企業管理學系碩士論文，1996。
- 張櫻淳，「醫院形象定位—以台北市六家醫學中心的家醫科就診民眾為例」，台灣大學公共衛生研究所碩士論文，1996。
- 陳怡潔，「醫療服務品質與顧客滿意之相關研究—彰化基督教醫院為例」，東海大學管理研究所碩士論文，1996。
- 黃俊智，「應用模擬技術探討某專科診所之門診預約掛號制度」，國立台灣大學醫療機構管理研究所碩士論文，1996。

- 楊朝欽，「醫院掛號作業效率之決策研究--等候理論之應用」，大葉工學院事業經營研究所碩士論文，1996。
- 廖慶榮，「作業研究」，第二版三刷，三民書局，台北市 2001。
- 劉敏玲，「急診病患對就醫之感受等候時間、實際等候時間與滿意度之相關研究」，國立臺灣大學醫療機構管理研究所碩士論文，2001。
- 劉偉文，「醫療行銷對病患滿意度之實證研究」，中正大學企業管理研究所碩士論文，1999。
- 盧昭文，「門診等候時間報告—壢新醫院為例」，2000。
- 謝崇耀 譯，「Oracle 資料庫設計」，初版，美商歐萊禮台灣分公司，台北市 2000。
- DB2 資訊月刊譯，Peter Gwynne 「用資料挖採技術 讓廢紙轉變成寶藏(1)」，DB2 資訊月刊 第五期 九月號，相關軟體學術合作計劃，交通大學計算機中心，1996。

英文文獻

- Anil. K. Jain. and Dubes Richard, Algorithms for Clustering Data. , Prentice-Hall
Advanced reference series. Prentice-Hall, Inc., Upper Saddle River, NJ. 1988.
- Berry J. A. Michael & Gordon Linoff , Data Mining Techniques . , New York: John
Wiley & Sons, Inc. 1997.
- Blanco White MJ, Pike MC. , Appointment systems in out-patients clinics and the
effect of patients unpunctuality., Med Care, 2: 133–45, 1964.
- Connolly M.Thomas & Begg E. Carolyn, , Database System – A practical approach to
design implementation and management , 2nd Edition, Addison Wesley, 1998.
- Dexter, Franklin MD, PhD. , Design of Appointment Systems for Preanesthesia
Evaluation Clinics to Minimize Patient Waiting Times: A Review of Computer
Simulation and Patient Survey Studies , International Anesthesia Research
Society 1999.
- Dubes, R.C. , How many clusters are best? An experiment. Pattern Recognition. ,
pp.645-663, Nov.1.1987.
- Fitzsimmons A. James , Service Management: Operations Strategy and Information
Technology , Part Three Managing Service Operations, 3rd Edition.
Boston:Irwin/McGraw Hill. 2002.
- Golfarelli, M., and Rizzi, S. , Designing the data warehouse: key steps and crucial
issues. , Journal of Computer Science and Information Management, vol. 2, n. 3,
1999.
- Lan H. Witten, Eibe Frank , Data Mining-Practical Machine Learning Tools and
Techniques with Java Implementations. , Chapter 6 Clustering pp.210-224
Morgan Kaufmann Publishers. 1999.
- Larson, Richard C., Perspectives on Queues: Social Justice and the Psychology of

Queuing, Operations Research, pp. 895-905 vol. 35, no. 6, November-December 1987.

Maister David, The Psychology of Waiting in Lines, Harvard Business Review, 1984.

Mitchell M.Tom., Machine Learning. , Chapter 6 The EM Algorithm pp.191-198. The McGraw-Hill Companies,Inc. 1997.

Rockhart JF, Herzog EL. , A predictive model for ambulatory patient service time. , Medical Care, 12: 512–9, 1974.

Tsai Yuh-Show, Paul H. King, Michael S. Higgins, Nonald Pierce, and Nimesh , An expert-guided decision tree construction strategy:An application in knowledge discovery with medical database , AMIA Annual fall symposium, pp.121-280, AMIA Nashville TN,1997.

Vissers J. , Selecting a suitable appointment system in an outpatient setting. , Medical Care, 17: 1207–20, 1979.

附 錄

附錄一 D63 醫師以 Index Mining 軟體所得到 EM 之結果

以下為取 D63 醫師以 Index Mining 軟體所得到 EM 之結果。

=== Run information ===

Scheme: EM -I 100 -N 5 -S 100 -M 1.0E-6

Relation: QueryResult

Instances: 1179

Attributes: 7

pt_type

date

shift

dr_v_time

age

sex

disease_code

Test mode: evaluate on training data

=== Clustering model (full training set) ===

EM

==

Number of clusters: 5

Cluster: 0 Prior probability: 0.3387

Attribute: pt_type

Discrete Estimator. Counts = | 3: 207.73 | 2: 178.73 | 1: 11.86 (Total = 398.31)

Attribute: date

Discrete Estimator. Counts = | 6: 0 | 3: 0 | 5: 0 | 1: 0 | 4: 175.38 | 2: 222.92 | 7: 0
(Total = 398.31)

Attribute: shift

Discrete Estimator. Counts = | 1: 0.01 | 2: 0.01 | 3: 398.3 (Total = 398.31)

Attribute: dr_v_time

Normal Distribution. Mean = 3.3477 StdDev = 1.7989

Attribute: age

Normal Distribution. Mean = 30.1084 StdDev = 17.4799

Attribute: sex

Discrete Estimator. Counts = | 1: 211.39 | 0: 186.93 (Total = 398.31)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 32.99 | 6809: 15.45 | 87340: 26.37 | 6819:
17.95 | 6829: 39.23 | 5289: 2 | 9049: 19.9 | 49390: 0 | 72741: 3 | 55090: 1.17 | 88003:
3.99 | 8830: 11.03 | 9490: 7.81 | 5679: 3.97 | 2164: 1 | 88100: 1 | 8820: 6 | 0781: 7.94 |
2169: 9.68 | 8798: 0 | 24200: 5 | 4661: 0 | 2740: 0 | 7890: 2.98 | 5409: 2.98 | 5212:
8.94 | 38010: 0 | 8290: 8.04 | 6119: 40.38 | 8940: 18.81 | 9194: 0 | 7856: 2 | 7823: 0.99
| 68111: 3 | 25001: 0 | 7806: 1 | 4659: 0 | 78650: 0.96 | 2140: 4.99 | 7294: 0 | 24290:
10.99 | 8840: 4 | 7842: 0 | 61171: 0 | 4556: 1 | 2393: 0 | 6110: 0 | 9983: 0 | 2893: 1 |
57140: 0.99 | 7039: 0 | 7840: 1 | 5950: 0 | 700: 5.04 | 7242: 0 | 2409: 1 | 1990: 0 | 2259:
0 | 7061: 1 | 57420: 0.98 | 73300: 1 | 7866: 0 | 94446: 0 | 53390: 0 | 81600: 3 | 81381: 0
| 2149: 1 | 193: 3 | 6929: 1.97 | 30002: 0 | 8930: 0 | 25000: 0 | 6272: 0 | 5609: 0 | 5759:
0 | 7062: 2.48 | 8920: 0 | V670: 0 | 4549: 2.67 | 7291: 4.03 | 4512: 1.97 | 3004: 0 |
37604: 0 | 4019: 0 | 25060: 0 | V048: 0 | 4351: 0 | 2449: 0 | 5921: 0 | 5368: 0 | 600: 0 |
7213: 1.96 | 1740: 1 | 1749: 0 | 49320: 0 | 1104: 0 | 37214: 0 | 34590: 0 | 8910: 1 | 438:
1 | 8921: 0 | 94405: 0 | 94800: 3 | 71940: 0 | 72889: 0 | 462: 0 | 82382: 0 | 8500: 1 |
9249: 0 | 7014: 3 | 78659: 0 | 4779: 1 | 53550: 0 | 9221: 0 | 94300: 0 | 2469: 0 | 9233: 0
| 2166: 0 | 72742: 1 | 61610: 0 | 605: 0 | 9243: 0 | 6149: 0 | 7893: 0 | 3670: 0 | 71690: 0
| 4539: 0 | 1519: 2.98 | 5641: 1.94 | 1539: 0 | 71590: 0 | 4660: 1 | 5999: 1 | 5282: 1 |
1109: 1 | 01200: 1 | 92401: 1 | 92411: 2 | 7245: 1 | 7071: 1 | 25040: 0.81 | 2102: 1 |
71500: 1 | 6980: 1 | 2400: 1 | 72210: 0.99 | 2749: 1 | 70400: 1 | 3540: 1 | 72252: 2 |
1579: 0.02 | 4738: 0 | 5789: 0 | 3469: 0 | 7094: 0 (Total = 398.31)

Cluster: 1 Prior probability: 0.2901

Attribute: pt_type

Discrete Estimator. Counts = | 3: 162.59 | 2: 154.18 | 1: 24.93 (Total = 341.7)

Attribute: date

Discrete Estimator. Counts = | 6: 152.19 | 3: 139.13 | 5: 9.73 | 1: 0 | 4: 0 | 2: 0 | 7:
40.66 (Total = 341.7)

Attribute: shift

Discrete Estimator. Counts = | 1: 341.67 | 2: 0.03 | 3: 0.01 (Total = 341.7)

Attribute: dr_v_time

Normal Distribution. Mean = 3.6181 StdDev = 1.8787

Attribute: age

Normal Distribution. Mean = 36.0586 StdDev = 20.1089

Attribute: sex

Discrete Estimator. Counts = | 1: 159.64 | 0: 182.07 (Total = 341.7)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 22.97 | 6809: 14.94 | 87340: 16.85 | 6819: 7 | 6829: 18.87 | 5289: 1.98 | 9049: 17.54 | 49390: 0.99 | 72741: 2 | 55090: 2.47 | 88003: 2 | 8830: 11.94 | 9490: 4.81 | 5679: 0.99 | 2164: 1 | 88100: 1 | 8820: 11.91 | 0781: 3.98 | 2169: 9.41 | 8798: 3.97 | 24200: 1.99 | 4661: 2 | 2740: 1 | 7890: 4.25 | 5409: 1 | 5212: 8.67 | 38010: 1 | 8290: 5.6 | 6119: 27.95 | 8940: 8.99 | 9194: 0.9 | 7856: 4.98 | 7823: 0.04 | 68111: 4.92 | 25001: 4.98 | 7806: 1.99 | 4659: 1.99 | 78650: 3 | 2140: 6.99 | 7294: 2.95 | 24290: 1 | 8840: 2 | 7842: 3 | 61171: 1 | 4556: 2.99 | 2393: 1.86 | 6110: 1 | 9983: 0 | 2893: 1 | 57140: 2.37 | 7039: 1 | 7840: 1 | 5950: 1 | 700: 5.99 | 7242: 2 | 2409: 3 | 1990: 0.74 | 2259: 1.31 | 7061: 0.24 | 57420: 1 | 73300: 1 | 7866: 0.99 | 94446: 0.02 | 53390: 0.89 | 81600: 2.97 | 81381: 1.99 | 2149: 1 | 193: 0.01 | 6929: 2 | 30002: 1 | 8930: 0.02 | 25000: 1 | 6272: 0.99 | 5609: 0 | 5759: 0.01 | 7062: 1.98 | 8920: 2 | V670: 1 | 4549: 2 | 7291: 5 | 4512: 2.99 | 3004: 1 | 37604: 1 | 4019: 2.97 | 25060: 1 | V048: 0.93 | 4351: 2.97 | 2449: 0.98 | 5921: 0.96 | 5368: 0.11 | 600: 1.99 | 7213: 2.08 | 1740: 0.71 | 1749: 1 | 49320: 1 | 1104: 0 | 37214: 0.89 | 34590: 0.99 | 8910: 0.99 | 438: 0.99 | 8921: 0.99 | 94405: 0 | 94800: 0 | 71940: 0 | 72889: 0 | 462: 0 | 82382: 0 | 8500: 0 | 9249: 0 | 7014: 0 | 78659: 0 | 4779: 0 | 53550: 0 | 9221: 0 | 94300: 0 | 2469: 0 | 9233: 0 | 2166: 0 | 72742: 0 | 61610: 0 | 605: 0 | 9243: 0 | 6149: 0 | 7893: 0 | 3670: 0 | 71690: 0 | 4539: 0 | 1519: 0 | 5641: 0 | 1539: 0 | 71590: 0 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 1 | 7245: 0 | 7071: 0 | 25040: 0 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0 | 2749: 0 | 70400: 0 | 3540: 0 | 72252: 0 | 1579: 0 | 4738: 0 | 5789: 1 | 3469: 1 | 7094: 0.99 (Total = 341.7)

Cluster: 2 Prior probability: 0.0381

Attribute: pt_type

Discrete Estimator. Counts = | 3: 21.77 | 2: 23.29 | 1: 1.74 (Total = 46.8)

Attribute: date

Discrete Estimator. Counts = | 6: 0.19 | 3: 0.25 | 5: 28.21 | 1: 17.86 | 4: 0.02 | 2: 0.03 | 7: 0.24 (Total = 46.8)

Attribute: shift

Discrete Estimator. Counts = | 1: 5.05 | 2: 41.7 | 3: 0.05 (Total = 46.8)

Attribute: dr_v_time

Normal Distribution. Mean = 11.865 StdDev = 4.0115

Attribute: age

Normal Distribution. Mean = 35.4439 StdDev = 17.8517

Attribute: sex

Discrete Estimator. Counts = | 1: 29.63 | 0: 17.17 (Total = 46.8)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 2.78 | 6809: 1.08 | 87340: 0.11 | 6819: 0.32 | 6829: 6.1 | 5289: 0 | 9049: 2.19 | 49390: 0 | 72741: 0 | 55090: 0.56 | 88003: 0.01 | 8830: 2.16 | 9490: 0.75 | 5679: 0.84 | 2164: 0 | 88100: 0 | 8820: 2.04 | 0781: 1.02 | 2169: 0.06 | 8798: 1.03 | 24200: 0.01 | 4661: 0.01 | 2740: 0.66 | 7890: 1.16 | 5409: 0.35 | 5212: 3.24 | 38010: 0 | 8290: 0.62 | 6119: 2.15 | 8940: 1.06 | 9194: 0.05 | 7856: 0.03 | 7823: 0.01 | 68111: 1.18 | 25001: 0.08 | 7806: 0 | 4659: 0.03 | 78650: 0 | 2140: 1.06 | 7294: 0.02 | 24290: 1.08 | 8840: 0.03 | 7842: 0.95 | 61171: 0.01 | 4556: 0.35 | 2393: 0.02 | 6110: 0.99 | 9983: 0.03 | 2893: 0 | 57140: 0.56 | 7039: 0.03 | 7840: 0 | 5950: 0 | 700: 0.1 | 7242: 0 | 2409: 0.24 | 1990: 0.01 | 2259: 0.01 | 7061: 0.01 | 57420: 0.98 | 73300: 0 | 7866: 0 | 94446: 0.01 | 53390: 0.34 | 81600: 1 | 81381: 0 | 2149: 0 | 193: 0.02 | 6929: 0.13 | 30002: 0.17 | 8930: 0.01 | 25000: 0.01 | 6272: 0.01 | 5609: 0 | 5759: 0.01 | 7062: 0.01 | 8920: 0.02 | V670: 0 | 4549: 0.01 | 7291: 0.01 | 4512: 0 | 3004: 0.05 | 37604: 0 | 4019: 2.11 | 25060: 0 | V048: 0.04 | 4351: 1 | 2449: 0.01 | 5921: 0 | 5368: 0.77 | 600: 1.01 | 7213: 0.01 | 1740: 0.02 | 1749: 0 | 49320: 0 | 1104: 0 | 37214: 0 | 34590: 0 | 8910: 0.01 | 438: 0 | 8921: 0 | 94405: 0 | 94800: 0.01 | 71940: 0 | 72889: 0.04 | 462: 0.02 | 82382: 0.01 | 8500: 0.01 | 9249: 0.01 | 7014: 0.01 | 78659: 0.02 | 4779: 0.45 | 53550: 0.01 | 9221: 0.01 | 94300: 0.02 | 2469: 0 | 9233: 0.01 | 2166: 0.01 | 72742: 0 | 61610: 0.85 | 605: 0.02 | 9243: 0.03 | 6149: 0.03 | 7893: 0.01 | 3670: 0.02 | 71690: 0.01 | 4539: 0.1 | 1519: 0 | 5641: 0 | 1539: 0.05 | 71590: 0.02 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 0 | 7245: 0 | 7071: 0 | 25040: 0 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0 | 2749: 0 | 70400: 0 | 3540: 0 | 72252: 0 | 1579: 0 | 4738: 0.04 | 5789: 0 | 3469: 0 | 7094: 0 (Total = 46.8)

Cluster: 3 Prior probability: 0.271

Attribute: pt_type

Discrete Estimator. Counts = | 3: 146.53 | 2: 155.03 | 1: 16.27 (Total = 317.83)

Attribute: date

Discrete Estimator. Counts = | 6: 0 | 3: 0 | 5: 126.69 | 1: 191.12 | 4: 0 | 2: 0 | 7: 0 (Total = 317.83)

Attribute: shift

Discrete Estimator. Counts = | 1: 0.59 | 2: 317.24 | 3: 0.01 (Total = 317.83)

Attribute: dr_v_time

Normal Distribution. Mean = 3.8263 StdDev = 2.1145

Attribute: age

Normal Distribution. Mean = 33.0233 StdDev = 20.4843

Attribute: sex

Discrete Estimator. Counts = | 1: 163.96 | 0: 153.87 (Total = 317.83)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 28.2 | 6809: 10.93 | 87340: 21.91 | 6819: 3.68 | 6829: 15.98 | 5289: 1 | 9049: 12.04 | 49390: 0 | 72741: 0 | 55090: 3.45 | 88003: 1.99 | 8830: 17.74 | 9490: 6.43 | 5679: 2.16 | 2164: 0 | 88100: 0 | 8820: 7 | 0781: 1.98 | 2169: 8.95 | 8798: 1.97 | 24200: 0.99 | 4661: 0.99 | 2740: 0.34 | 7890: 6.87 | 5409: 4.66 | 5212: 6.82 | 38010: 0 | 8290: 4.4 | 6119: 28.8 | 8940: 3.94 | 9194: 0.96 | 7856: 2.98 | 7823: 0 | 68111: 1.87 | 25001: 0.92 | 7806: 0 | 4659: 1 | 78650: 0 | 2140: 1.94 | 7294: 1.03 | 24290: 3.95 | 8840: 4.98 | 7842: 4.05 | 61171: 0.99 | 4556: 2.84 | 2393: 0.99 | 6110: 1.01 | 9983: 1.98 | 2893: 0 | 57140: 1.02 | 7039: 1.97 | 7840: 0 | 5950: 0 | 700: 4.9 | 7242: 0 | 2409: 5.76 | 1990: 0 | 2259: 1 | 7061: 0 | 57420: 0.02 | 73300: 0 | 7866: 0 | 94446: 0 | 53390: 1.79 | 81600: 1 | 81381: 0 | 2149: 0 | 193: 0.99 | 6929: 2.87 | 30002: 0.83 | 8930: 1 | 25000: 0.99 | 6272: 0.99 | 5609: 0 | 5759: 0 | 7062: 1.99 | 8920: 0.98 | V670: 0 | 4549: 0.99 | 7291: 0.99 | 4512: 0 | 3004: 0.95 | 37604: 0 | 4019: 2.89 | 25060: 0 | V048: 0.03 | 4351: 0 | 2449: 0.99 | 5921: 0 | 5368: 1.01 | 600: 0.99 | 7213: 1 | 1740: 1.98 | 1749: 0 | 49320: 0 | 1104: 0 | 37214: 0 | 34590: 0 | 8910: 1.99 | 438: 0 | 8921: 0 | 94405: 1 | 94800: 1.99 | 71940: 1 | 72889: 0.96 | 462: 0.98 | 82382: 0.99 | 8500: 0.99 | 9249: 1.99 | 7014: 0.98 | 78659: 2.98 | 4779: 0.55 | 53550: 1.99 | 9221: 0.99 | 94300: 0.98 | 2469: 1 | 9233: 0.99 | 2166: 0.99 | 72742: 1 | 61610: 0.15 | 605: 0.98 | 9243: 0.97 | 6149: 0.97 | 7893: 1.99 | 3670: 1.98 | 71690: 0.99 | 4539: 3.9 | 1519: 1 | 5641: 1 | 1539: 3.95 | 71590: 0.98 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 0 | 7245: 0 | 7071: 0 | 25040: 0 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0 | 2749: 0 | 70400: 0 | 3540: 0 | 72252: 0 | 1579: 0 | 4738: 0 | 5789: 0 | 3469: 0 | 7094: 0 (Total = 317.83)

Cluster: 4 Prior probability: 0.0621

Attribute: pt_type

Discrete Estimator. Counts = | 3: 48.38 | 2: 23.77 | 1: 2.19 (Total = 74.35)

Attribute: date

Discrete Estimator. Counts = | 6: 14.62 | 3: 18.61 | 5: 0.37 | 1: 0.01 | 4: 20.59 | 2: 18.05 | 7: 2.09 (Total = 74.35)

Attribute: shift

Discrete Estimator. Counts = | 1: 35.68 | 2: 0.02 | 3: 38.64 (Total = 74.35)

Attribute: dr_v_time

Normal Distribution. Mean = 11.4713 StdDev = 3.2281

Attribute: age

Normal Distribution. Mean = 41.66 StdDev = 21.1914

Attribute: sex

Discrete Estimator. Counts = | 1: 49.38 | 0: 24.96 (Total = 74.35)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 1.06 | 6809: 3.61 | 87340: 3.76 | 6819: 0.06 | 6829: 5.81 | 5289: 0.02 | 9049: 3.34 | 49390: 0 | 72741: 0 | 55090: 2.35 | 88003: 0.01 | 8830: 2.13 | 9490: 0.19 | 5679: 0.04 | 2164: 0 | 88100: 0 | 8820: 0.05 | 0781: 0.08 | 2169: 3.89 | 8798: 0.02 | 24200: 0.01 | 4661: 0 | 2740: 0 | 7890: 0.74 | 5409: 2.01 | 5212: 2.33 | 38010: 0 | 8290: 1.35 | 6119: 3.72 | 8940: 1.2 | 9194: 0.09 | 7856: 1.01 | 7823: 0.95 | 68111: 0.03 | 25001: 0.02 | 7806: 0.01 | 4659: 0.98 | 78650: 0.04 | 2140: 1.02 | 7294: 0 | 24290: 1.98 | 8840: 0.99 | 7842: 0 | 61171: 0 | 4556: 0.82 | 2393: 2.13 | 6110: 0 | 9983: 0.99 | 2893: 0 | 57140: 0.06 | 7039: 0 | 7840: 0 | 5950: 0 | 700: 0.97 | 7242: 0 | 2409: 0 | 1990: 0.25 | 2259: 0.68 | 7061: 0.76 | 57420: 0.02 | 73300: 0 | 7866: 0.01 | 94446: 0.97 | 53390: 1.98 | 81600: 0.03 | 81381: 0.01 | 2149: 0 | 193: 0.98 | 6929: 0.03 | 30002: 0 | 8930: 0.97 | 25000: 0 | 6272: 0 | 5609: 0.99 | 5759: 0.98 | 7062: 0.55 | 8920: 0 | V670: 0 | 4549: 0.33 | 7291: 0.98 | 4512: 0.04 | 3004: 0 | 37604: 0 | 4019: 0.03 | 25060: 0 | V048: 0 | 4351: 1.02 | 2449: 1.02 | 5921: 0.03 | 5368: 0.11 | 600: 0.01 | 7213: 0.96 | 1740: 0.29 | 1749: 0 | 49320: 0 | 1104: 1 | 37214: 0.11 | 34590: 0.01 | 8910: 0.01 | 438: 1.01 | 8921: 0.01 | 94405: 0 | 94800: 0 | 71940: 0 | 72889: 0 | 462: 0 | 82382: 0 | 8500: 0 | 9249: 0 | 7014: 0 | 78659: 0 | 4779: 0 | 53550: 0 | 9221: 0 | 94300: 0 | 2469: 0 | 9233: 0 | 2166: 0 | 72742: 0 | 61610: 0 | 605: 0 | 9243: 0 | 6149: 0 | 7893: 0 | 3670: 0 | 71690: 0 | 4539: 0 | 1519: 0.02 | 5641: 2.06 | 1539: 0 | 71590: 0 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 0 | 7245: 0 | 7071: 0 | 25040: 0.19 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0.01 | 2749: 0 | 70400: 0 | 3540: 1 | 72252: 0 | 1579: 3.98 | 4738: 0.95 | 5789: 0 | 3469: 0 | 7094: 0.01 (Total = 74.35)

=== Evaluation on training set ===

EM

==

Number of clusters: 5

Cluster: 0 Prior probability: 0.3387

Attribute: pt_type

Discrete Estimator. Counts = | 3: 207.73 | 2: 178.73 | 1: 11.86 (Total = 398.31)

Attribute: date

Discrete Estimator. Counts = | 6: 0 | 3: 0 | 5: 0 | 1: 0 | 4: 175.38 | 2: 222.92 | 7: 0

(Total = 398.31)

Attribute: shift

Discrete Estimator. Counts = | 1: 0.01 | 2: 0.01 | 3: 398.3 (Total = 398.31)

Attribute: dr_v_time

Normal Distribution. Mean = 3.3477 StdDev = 1.7989

Attribute: age

Normal Distribution. Mean = 30.1084 StdDev = 17.4799

Attribute: sex

Discrete Estimator. Counts = | 1: 211.39 | 0: 186.93 (Total = 398.31)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 32.99 | 6809: 15.45 | 87340: 26.37 | 6819:
17.95 | 6829: 39.23 | 5289: 2 | 9049: 19.9 | 49390: 0 | 72741: 3 | 55090: 1.17 | 88003:
3.99 | 8830: 11.03 | 9490: 7.81 | 5679: 3.97 | 2164: 1 | 88100: 1 | 8820: 6 | 0781: 7.94 |
2169: 9.68 | 8798: 0 | 24200: 5 | 4661: 0 | 2740: 0 | 7890: 2.98 | 5409: 2.98 | 5212:
8.94 | 38010: 0 | 8290: 8.04 | 6119: 40.38 | 8940: 18.81 | 9194: 0 | 7856: 2 | 7823: 0.99
| 68111: 3 | 25001: 0 | 7806: 1 | 4659: 0 | 78650: 0.96 | 2140: 4.99 | 7294: 0 | 24290:
10.99 | 8840: 4 | 7842: 0 | 61171: 0 | 4556: 1 | 2393: 0 | 6110: 0 | 9983: 0 | 2893: 1 |
57140: 0.99 | 7039: 0 | 7840: 1 | 5950: 0 | 700: 5.04 | 7242: 0 | 2409: 1 | 1990: 0 | 2259:
0 | 7061: 1 | 57420: 0.98 | 73300: 1 | 7866: 0 | 94446: 0 | 53390: 0 | 81600: 3 | 81381: 0
| 2149: 1 | 193: 3 | 6929: 1.97 | 30002: 0 | 8930: 0 | 25000: 0 | 6272: 0 | 5609: 0 | 5759:
0 | 7062: 2.48 | 8920: 0 | V670: 0 | 4549: 2.67 | 7291: 4.03 | 4512: 1.97 | 3004: 0 |
37604: 0 | 4019: 0 | 25060: 0 | V048: 0 | 4351: 0 | 2449: 0 | 5921: 0 | 5368: 0 | 600: 0 |
7213: 1.96 | 1740: 1 | 1749: 0 | 49320: 0 | 1104: 0 | 37214: 0 | 34590: 0 | 8910: 1 | 438:
1 | 8921: 0 | 94405: 0 | 94800: 3 | 71940: 0 | 72889: 0 | 462: 0 | 82382: 0 | 8500: 1 |
9249: 0 | 7014: 3 | 78659: 0 | 4779: 1 | 53550: 0 | 9221: 0 | 94300: 0 | 2469: 0 | 9233: 0
| 2166: 0 | 72742: 1 | 61610: 0 | 605: 0 | 9243: 0 | 6149: 0 | 7893: 0 | 3670: 0 | 71690: 0
| 4539: 0 | 1519: 2.98 | 5641: 1.94 | 1539: 0 | 71590: 0 | 4660: 1 | 5999: 1 | 5282: 1 |
1109: 1 | 01200: 1 | 92401: 1 | 92411: 2 | 7245: 1 | 7071: 1 | 25040: 0.81 | 2102: 1 |
71500: 1 | 6980: 1 | 2400: 1 | 72210: 0.99 | 2749: 1 | 70400: 1 | 3540: 1 | 72252: 2 |
1579: 0.02 | 4738: 0 | 5789: 0 | 3469: 0 | 7094: 0 (Total = 398.31)

Cluster: 1 Prior probability: 0.2901

Attribute: pt_type

Discrete Estimator. Counts = | 3: 162.59 | 2: 154.18 | 1: 24.93 (Total = 341.7)

Attribute: date

Discrete Estimator. Counts = | 6: 152.19 | 3: 139.13 | 5: 9.73 | 1: 0 | 4: 0 | 2: 0 | 7:
40.66 (Total = 341.7)

Attribute: shift

Discrete Estimator. Counts = | 1: 341.67 | 2: 0.03 | 3: 0.01 (Total = 341.7)

Attribute: dr_v_time

Normal Distribution. Mean = 3.6181 StdDev = 1.8787

Attribute: age

Normal Distribution. Mean = 36.0586 StdDev = 20.1089

Attribute: sex

Discrete Estimator. Counts = | 1: 159.64 | 0: 182.07 (Total = 341.7)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 22.97 | 6809: 14.94 | 87340: 16.85 | 6819: 7 | 6829: 18.87 | 5289: 1.98 | 9049: 17.54 | 49390: 0.99 | 72741: 2 | 55090: 2.47 | 88003: 2 | 8830: 11.94 | 9490: 4.81 | 5679: 0.99 | 2164: 1 | 88100: 1 | 8820: 11.91 | 0781: 3.98 | 2169: 9.41 | 8798: 3.97 | 24200: 1.99 | 4661: 2 | 2740: 1 | 7890: 4.25 | 5409: 1 | 5212: 8.67 | 38010: 1 | 8290: 5.6 | 6119: 27.95 | 8940: 8.99 | 9194: 0.9 | 7856: 4.98 | 7823: 0.04 | 68111: 4.92 | 25001: 4.98 | 7806: 1.99 | 4659: 1.99 | 78650: 3 | 2140: 6.99 | 7294: 2.95 | 24290: 1 | 8840: 2 | 7842: 3 | 61171: 1 | 4556: 2.99 | 2393: 1.86 | 6110: 1 | 9983: 0 | 2893: 1 | 57140: 2.37 | 7039: 1 | 7840: 1 | 5950: 1 | 700: 5.99 | 7242: 2 | 2409: 3 | 1990: 0.74 | 2259: 1.31 | 7061: 0.24 | 57420: 1 | 73300: 1 | 7866: 0.99 | 94446: 0.02 | 53390: 0.89 | 81600: 2.97 | 81381: 1.99 | 2149: 1 | 193: 0.01 | 6929: 2 | 30002: 1 | 8930: 0.02 | 25000: 1 | 6272: 0.99 | 5609: 0 | 5759: 0.01 | 7062: 1.98 | 8920: 2 | V670: 1 | 4549: 2 | 7291: 5 | 4512: 2.99 | 3004: 1 | 37604: 1 | 4019: 2.97 | 25060: 1 | V048: 0.93 | 4351: 2.97 | 2449: 0.98 | 5921: 0.96 | 5368: 0.11 | 600: 1.99 | 7213: 2.08 | 1740: 0.71 | 1749: 1 | 49320: 1 | 1104: 0 | 37214: 0.89 | 34590: 0.99 | 8910: 0.99 | 438: 0.99 | 8921: 0.99 | 94405: 0 | 94800: 0 | 71940: 0 | 72889: 0 | 462: 0 | 82382: 0 | 8500: 0 | 9249: 0 | 7014: 0 | 78659: 0 | 4779: 0 | 53550: 0 | 9221: 0 | 94300: 0 | 2469: 0 | 9233: 0 | 2166: 0 | 72742: 0 | 61610: 0 | 605: 0 | 9243: 0 | 6149: 0 | 7893: 0 | 3670: 0 | 71690: 0 | 4539: 0 | 1519: 0 | 5641: 0 | 1539: 0 | 71590: 0 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 1 | 7245: 0 | 7071: 0 | 25040: 0 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0 | 2749: 0 | 70400: 0 | 3540: 0 | 72252: 0 | 1579: 0 | 4738: 0 | 5789: 1 | 3469: 1 | 7094: 0.99 (Total = 341.7)

Cluster: 2 Prior probability: 0.0381

Attribute: pt_type

Discrete Estimator. Counts = | 3: 21.77 | 2: 23.29 | 1: 1.74 (Total = 46.8)

Attribute: date

Discrete Estimator. Counts = | 6: 0.19 | 3: 0.25 | 5: 28.21 | 1: 17.86 | 4: 0.02 | 2: 0.03 | 7: 0.24 (Total = 46.8)

Attribute: shift

Discrete Estimator. Counts = | 1: 5.05 | 2: 41.7 | 3: 0.05 (Total = 46.8)

Attribute: dr_v_time

Normal Distribution. Mean = 11.865 StdDev = 4.0115

Attribute: age

Normal Distribution. Mean = 35.4439 StdDev = 17.8517

Attribute: sex

Discrete Estimator. Counts = | 1: 29.63 | 0: 17.17 (Total = 46.8)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 2.78 | 6809: 1.08 | 87340: 0.11 | 6819: 0.32 | 6829: 6.1 | 5289: 0 | 9049: 2.19 | 49390: 0 | 72741: 0 | 55090: 0.56 | 88003: 0.01 | 8830: 2.16 | 9490: 0.75 | 5679: 0.84 | 2164: 0 | 88100: 0 | 8820: 2.04 | 0781: 1.02 | 2169: 0.06 | 8798: 1.03 | 24200: 0.01 | 4661: 0.01 | 2740: 0.66 | 7890: 1.16 | 5409: 0.35 | 5212: 3.24 | 38010: 0 | 8290: 0.62 | 6119: 2.15 | 8940: 1.06 | 9194: 0.05 | 7856: 0.03 | 7823: 0.01 | 68111: 1.18 | 25001: 0.08 | 7806: 0 | 4659: 0.03 | 78650: 0 | 2140: 1.06 | 7294: 0.02 | 24290: 1.08 | 8840: 0.03 | 7842: 0.95 | 61171: 0.01 | 4556: 0.35 | 2393: 0.02 | 6110: 0.99 | 9983: 0.03 | 2893: 0 | 57140: 0.56 | 7039: 0.03 | 7840: 0 | 5950: 0 | 700: 0.1 | 7242: 0 | 2409: 0.24 | 1990: 0.01 | 2259: 0.01 | 7061: 0.01 | 57420: 0.98 | 73300: 0 | 7866: 0 | 94446: 0.01 | 53390: 0.34 | 81600: 1 | 81381: 0 | 2149: 0 | 193: 0.02 | 6929: 0.13 | 30002: 0.17 | 8930: 0.01 | 25000: 0.01 | 6272: 0.01 | 5609: 0 | 5759: 0.01 | 7062: 0.01 | 8920: 0.02 | V670: 0 | 4549: 0.01 | 7291: 0.01 | 4512: 0 | 3004: 0.05 | 37604: 0 | 4019: 2.11 | 25060: 0 | V048: 0.04 | 4351: 1 | 2449: 0.01 | 5921: 0 | 5368: 0.77 | 600: 1.01 | 7213: 0.01 | 1740: 0.02 | 1749: 0 | 49320: 0 | 1104: 0 | 37214: 0 | 34590: 0 | 8910: 0.01 | 438: 0 | 8921: 0 | 94405: 0 | 94800: 0.01 | 71940: 0 | 72889: 0.04 | 462: 0.02 | 82382: 0.01 | 8500: 0.01 | 9249: 0.01 | 7014: 0.01 | 78659: 0.02 | 4779: 0.45 | 53550: 0.01 | 9221: 0.01 | 94300: 0.02 | 2469: 0 | 9233: 0.01 | 2166: 0.01 | 72742: 0 | 61610: 0.85 | 605: 0.02 | 9243: 0.03 | 6149: 0.03 | 7893: 0.01 | 3670: 0.02 | 71690: 0.01 | 4539: 0.1 | 1519: 0 | 5641: 0 | 1539: 0.05 | 71590: 0.02 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 0 | 7245: 0 | 7071: 0 | 25040: 0 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0 | 2749: 0 | 70400: 0 | 3540: 0 | 72252: 0 | 1579: 0 | 4738: 0.04 | 5789: 0 | 3469: 0 | 7094: 0 (Total = 46.8)

Cluster: 3 Prior probability: 0.271

Attribute: pt_type

Discrete Estimator. Counts = | 3: 146.53 | 2: 155.03 | 1: 16.27 (Total = 317.83)

Attribute: date

Discrete Estimator. Counts = | 6: 0 | 3: 0 | 5: 126.69 | 1: 191.12 | 4: 0 | 2: 0 | 7: 0

(Total = 317.83)

Attribute: shift

Discrete Estimator. Counts = | 1: 0.59 | 2: 317.24 | 3: 0.01 (Total = 317.83)

Attribute: dr_v_time

Normal Distribution. Mean = 3.8263 StdDev = 2.1145

Attribute: age

Normal Distribution. Mean = 33.0233 StdDev = 20.4843

Attribute: sex

Discrete Estimator. Counts = | 1: 163.96 | 0: 153.87 (Total = 317.83)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 28.2 | 6809: 10.93 | 87340: 21.91 | 6819: 3.68 | 6829: 15.98 | 5289: 1 | 9049: 12.04 | 49390: 0 | 72741: 0 | 55090: 3.45 | 88003: 1.99 | 8830: 17.74 | 9490: 6.43 | 5679: 2.16 | 2164: 0 | 88100: 0 | 8820: 7 | 0781: 1.98 | 2169: 8.95 | 8798: 1.97 | 24200: 0.99 | 4661: 0.99 | 2740: 0.34 | 7890: 6.87 | 5409: 4.66 | 5212: 6.82 | 38010: 0 | 8290: 4.4 | 6119: 28.8 | 8940: 3.94 | 9194: 0.96 | 7856: 2.98 | 7823: 0 | 68111: 1.87 | 25001: 0.92 | 7806: 0 | 4659: 1 | 78650: 0 | 2140: 1.94 | 7294: 1.03 | 24290: 3.95 | 8840: 4.98 | 7842: 4.05 | 61171: 0.99 | 4556: 2.84 | 2393: 0.99 | 6110: 1.01 | 9983: 1.98 | 2893: 0 | 57140: 1.02 | 7039: 1.97 | 7840: 0 | 5950: 0 | 700: 4.9 | 7242: 0 | 2409: 5.76 | 1990: 0 | 2259: 1 | 7061: 0 | 57420: 0.02 | 73300: 0 | 7866: 0 | 94446: 0 | 53390: 1.79 | 81600: 1 | 81381: 0 | 2149: 0 | 193: 0.99 | 6929: 2.87 | 30002: 0.83 | 8930: 1 | 25000: 0.99 | 6272: 0.99 | 5609: 0 | 5759: 0 | 7062: 1.99 | 8920: 0.98 | V670: 0 | 4549: 0.99 | 7291: 0.99 | 4512: 0 | 3004: 0.95 | 37604: 0 | 4019: 2.89 | 25060: 0 | V048: 0.03 | 4351: 0 | 2449: 0.99 | 5921: 0 | 5368: 1.01 | 600: 0.99 | 7213: 1 | 1740: 1.98 | 1749: 0 | 49320: 0 | 1104: 0 | 37214: 0 | 34590: 0 | 8910: 1.99 | 438: 0 | 8921: 0 | 94405: 1 | 94800: 1.99 | 71940: 1 | 72889: 0.96 | 462: 0.98 | 82382: 0.99 | 8500: 0.99 | 9249: 1.99 | 7014: 0.98 | 78659: 2.98 | 4779: 0.55 | 53550: 1.99 | 9221: 0.99 | 94300: 0.98 | 2469: 1 | 9233: 0.99 | 2166: 0.99 | 72742: 1 | 61610: 0.15 | 605: 0.98 | 9243: 0.97 | 6149: 0.97 | 7893: 1.99 | 3670: 1.98 | 71690: 0.99 | 4539: 3.9 | 1519: 1 | 5641: 1 | 1539: 3.95 | 71590: 0.98 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 0 | 7245: 0 | 7071: 0 | 25040: 0 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0 | 2749: 0 | 70400: 0 | 3540: 0 | 72252: 0 | 1579: 0 | 4738: 0 | 5789: 0 | 3469: 0 | 7094: 0 (Total = 317.83)

Cluster: 4 Prior probability: 0.0621

Attribute: pt_type

Discrete Estimator. Counts = | 3: 48.38 | 2: 23.77 | 1: 2.19 (Total = 74.35)

Attribute: date

Discrete Estimator. Counts = | 6: 14.62 | 3: 18.61 | 5: 0.37 | 1: 0.01 | 4: 20.59 | 2: 18.05 | 7: 2.09 (Total = 74.35)

Attribute: shift

Discrete Estimator. Counts = | 1: 35.68 | 2: 0.02 | 3: 38.64 (Total = 74.35)

Attribute: dr_v_time

Normal Distribution. Mean = 11.4713 StdDev = 3.2281

Attribute: age

Normal Distribution. Mean = 41.66 StdDev = 21.1914

Attribute: sex

Discrete Estimator. Counts = | 1: 49.38 | 0: 24.96 (Total = 74.35)

Attribute: disease_code

Discrete Estimator. Counts = | 8730: 1.06 | 6809: 3.61 | 87340: 3.76 | 6819: 0.06 | 6829: 5.81 | 5289: 0.02 | 9049: 3.34 | 49390: 0 | 72741: 0 | 55090: 2.35 | 88003: 0.01 | 8830: 2.13 | 9490: 0.19 | 5679: 0.04 | 2164: 0 | 88100: 0 | 8820: 0.05 | 0781: 0.08 | 2169: 3.89 | 8798: 0.02 | 24200: 0.01 | 4661: 0 | 2740: 0 | 7890: 0.74 | 5409: 2.01 | 5212: 2.33 | 38010: 0 | 8290: 1.35 | 6119: 3.72 | 8940: 1.2 | 9194: 0.09 | 7856: 1.01 | 7823: 0.95 | 68111: 0.03 | 25001: 0.02 | 7806: 0.01 | 4659: 0.98 | 78650: 0.04 | 2140: 1.02 | 7294: 0 | 24290: 1.98 | 8840: 0.99 | 7842: 0 | 61171: 0 | 4556: 0.82 | 2393: 2.13 | 6110: 0 | 9983: 0.99 | 2893: 0 | 57140: 0.06 | 7039: 0 | 7840: 0 | 5950: 0 | 700: 0.97 | 7242: 0 | 2409: 0 | 1990: 0.25 | 2259: 0.68 | 7061: 0.76 | 57420: 0.02 | 73300: 0 | 7866: 0.01 | 94446: 0.97 | 53390: 1.98 | 81600: 0.03 | 81381: 0.01 | 2149: 0 | 193: 0.98 | 6929: 0.03 | 30002: 0 | 8930: 0.97 | 25000: 0 | 6272: 0 | 5609: 0.99 | 5759: 0.98 | 7062: 0.55 | 8920: 0 | V670: 0 | 4549: 0.33 | 7291: 0.98 | 4512: 0.04 | 3004: 0 | 37604: 0 | 4019: 0.03 | 25060: 0 | V048: 0 | 4351: 1.02 | 2449: 1.02 | 5921: 0.03 | 5368: 0.11 | 600: 0.01 | 7213: 0.96 | 1740: 0.29 | 1749: 0 | 49320: 0 | 1104: 1 | 37214: 0.11 | 34590: 0.01 | 8910: 0.01 | 438: 1.01 | 8921: 0.01 | 94405: 0 | 94800: 0 | 71940: 0 | 72889: 0 | 462: 0 | 82382: 0 | 8500: 0 | 9249: 0 | 7014: 0 | 78659: 0 | 4779: 0 | 53550: 0 | 9221: 0 | 94300: 0 | 2469: 0 | 9233: 0 | 2166: 0 | 72742: 0 | 61610: 0 | 605: 0 | 9243: 0 | 6149: 0 | 7893: 0 | 3670: 0 | 71690: 0 | 4539: 0 | 1519: 0.02 | 5641: 2.06 | 1539: 0 | 71590: 0 | 4660: 0 | 5999: 0 | 5282: 0 | 1109: 0 | 01200: 0 | 92401: 0 | 92411: 0 | 7245: 0 | 7071: 0 | 25040: 0.19 | 2102: 0 | 71500: 0 | 6980: 0 | 2400: 0 | 72210: 0.01 | 2749: 0 | 70400: 0 | 3540: 1 | 72252: 0 | 1579: 3.98 | 4738: 0.95 | 5789: 0 | 3469: 0 | 7094: 0.01 (Total = 74.35)

Clustered Instances

0 437 (37%)

1 382 (32%)

2 360 (31%)

Log likelihood: -8.44279