

## 醫院門診資料的知識發現-以虎尾若瑟醫院為例

蔡鴻旭  
國立虎尾技術學院資訊  
管理系副教授兼系主任  
thh@sunws.nhit.edu.tw

李姿儀  
南華大學資訊管理學系碩士  
班研究生  
g8112014@mail1.nhu.edu.tw

鍾志明  
南華大學資訊管理學系助  
理教授  
cmchung@mail.nhu.edu.tw

## 摘要

本論文針對醫院門診的病歷資料庫，利用 Apriori 演算法之資料探勘技術，探勘出疾病間之關聯性知識法則。研究中以 816,558 筆就診記錄進行分析，並設定支持度為 0.3%，可信度為 90%，研究結果共得到 2 則有用的知識法則。此研究成果，可做為醫師診斷疾病之參考，並且提供一般民眾在健康保健方面能達到疾病預警之功效，同時也作為醫務管理之重要參考。

**關鍵字：**資料探勘、關聯法則、醫院門診資料、疾病預警、醫務管理

## 壹、緒論

就一個區域醫院來說，平均每天約有一千人次的病患就診，長期所累積下來的龐大資料中，蘊藏了許多隱而未知的有用知識。而資訊科技的發達，不僅在運算能力方面更為強大，儲存設備容量更形擴大，同時所提供的資訊軟體技術也更加豐富。因此藉由資訊科技的資料探勘技術，將病歷資料庫中有用的知識探索出來，以有效地找出具可參考性的醫學或醫務管理等方面的知識。

## 一、知識發現之簡述

根據[2]所定義的，知識發現是從資料中有意義地抽取出內隱的

(implicit)、先前未知的 (previously unknown) 和潛在性有用 (potentially useful) 的資訊。

資料探勘是屬於整個知識發現 (Knowledge Discovery in Database, KDD) [3,9,10,17]過程的重要處理方法之一，如圖 1 所示。



圖 1：KDD 流程圖

整個的知識發現過程分為三大步驟，闡述如下：

## 1. 資料淨化

檢視篩選出來的資料，刪除或修正錯誤的資料，或加入該再補充的資料內容，以達到資料一致性及完整性的要求。另外為了簡化運算的複雜度，將資料透過適當的轉換機制以重新編碼。最後將資料整理成能以資料探勘工具直接處理的型態，則資料淨化的工作於是完成。

## 2. 資料探勘

常見的資料探勘所要解決的問題有以下幾類：

A. 關聯性 (association) 的推導：零售業者就消費者的每筆消費記錄中，找出所有被購買產品之間的關

聯性，例如：當甲顧客買了產品 A 又買了產品 B 後，必定會再購買產品 C，即  $A \text{ and } B \rightarrow C$  就表示出產品 A、B、C 之間的關聯性 [1, 17]。

- B. 序列性 (sequences) 關聯法則推導：此問題的資料特性是與時間因素息息相關的，可分為時間序列樣型 (patterns within time series) 與循序樣型 (sequential patterns) 兩大類。時間序列樣型是分析在某時間區段內所發生的事件，是否也會在其他相同的時段內發生，所著重的是去偵查出在時間區段內所發生事件之相似性[1,17]。例如：如果 AT&T 的股票連續兩天上漲，而 DEC 的股票在這兩天並沒有下跌的話，那麼 IBM 的股票隔天就會上漲[4]。而循序樣型是考慮時間先後發生的事件狀況，去做資料關聯性分析[1]。例如：病人因動脈瘤而接受心導管分流外科手術，在其術後一年內產生高血尿的症狀，則他將在往後的 18 月內可能會面臨腎臟之衰竭[17]。即將時間的因素加入考慮，抽取序列性關聯法則，以解決牽涉到和時間推移相關的決策需求，就是序列性關聯法則推導這一類的問題[1, 4, 17]。
- C. 分類 (classification)：此問題是建立資料分類的法則，利用督導式的機器學習方法，配合一組訓練型樣 (training patterns)，建立一套資料分類法。而訓練型樣內含有類別的個數及屬性。例如：銀行想將借貸信用良好與借貸信用不良的客戶區分出來，因此設定達到 25% 的償債責任為一個門檻值，凡超過者是為借貸信用良好這一群；反之其餘者則為信用不良之另一群[1, 11, 17]。
- D. 群集 (clustering)：與分類問題相同，其主要不同之處是藉由非督導式的機器學習，將具有高度相似度的資料分類成一類。也就是，在資料的分群過程中，並未事先定義個群組的屬性與個數[1, 12]。

一般用來解決資料探勘問題的工具或方法則有關聯法則 (association rules)、決策樹 (decision trees)、類神經網路 (neural networks)、基因演算法 (genetic algorithms)、模糊推論系統 (fuzzy inference systems) 等工具來探勘有用的知識[1, 8, 9, 11, 17]。本論文即是利用 association rule 的推導演算法去發現門診疾病資料之間的關聯性。

### 3. 探勘資料釋義

探勘出來的資料常常是以簡單的符號或代碼所表示，因此為了讓人類更了解其中之奧妙，就必須將其轉為人類所能理解的自然語言並且輔以合理的解釋。最後再將這些獲得的知識整理好並放入知識庫中，提供管理人士參考。本論文將門診資料的關聯性法則抽取出來，提供醫師診斷、醫務管理及疾病預警等三方面之參考。

## 二、若瑟醫院簡述

位於雲林縣虎尾鎮的若瑟醫院，是於西元 1955 年由嘉義地區牛會卿主教，為醫療資源貧乏且醫療保健工作落後的雲林地區而興建，以提供當地區民有更良好的醫療品質。

在院長 松喬神父與副院長 畢耀遠神父的努力之下，若瑟醫院由早期的幾十個病床增至目前上百個病床，而預計於 2001 年新醫療大樓完成之時，病床將多達 700 多床。而院內之服務項目提供內科、外科、婦產科及小兒科等專科門診及住院服務，及設置洗腎中心加護病房與安寧病房等，為雲林地區之民眾提供優良的臨床優良救護。

院長 松喬神父與副院長 畢耀遠神父，也分別於 1993 年及 1995 年獲得第三屆與第五屆醫療奉獻獎的肯定。而若瑟醫院本身更於 89 年度醫院評鑑結果中，升等為區域醫院[19]。

## 貳、關聯法則之闡釋

關聯法則最早由 Agrawal et al. 於 1993 年所提出的[4]，並隨後於 1994 及 1995 年提出 Apriori 演算法[1, 6, 7]運用在關聯法則的推導。本論文採用 Apriori 演算法抽取門診資料知識，相關符號定義及其演算法簡述如下：

### 一、關聯法則之定義

令  $I = \{i_1, i_2, \dots, i_m\}$  是所有相異項目 (items) 所成的集合。假設資料庫  $D$  有  $n$  個交易記錄，即  $T = \{t_1, t_2, \dots, t_n\}$ ，而每筆交易記錄， $t_p$ ，都是由一部份的項目所組成之集合，即  $t_p \subseteq I$ ，其中  $p = 1, 2, \dots, n$ 。某一集合  $X$ ，若  $X \subseteq I$ ，則稱  $X$  為項目組 (itemset)，若此  $X$  包含  $k$  個項目，則稱  $X$  為  $k$ -項目組 ( $k$ -itemset)。

關聯法則之表示法如下所示： $X \rightarrow Y$ ，此時  $X, Y \subseteq I$  且  $X \cap Y = \emptyset$ 。藉由兩個指標值：支持度 (support) 及可靠度 (confidence) 來推導此關聯法則。 $X \rightarrow Y$  支持度定義如下： $support(X \cup Y) = \#(X \cup Y) / n$ ， $\#(X \cup Y)$  表示  $D$  中所有包含  $X \cup Y$  項目組之交易記錄個數， $0 \leq support(X \cup Y) \leq 1$ 。可靠度定義如下： $support(X \cup Y) / support(X)$ ，即此關聯法則  $X \rightarrow Y$  可信的程度，也就是說，在出現  $X$  情況的交易記錄個數，和  $X$  與  $Y$  同時出現 (即  $X \cup Y$ ) 的交易記錄個數的比率。而當支持度與可信度大於或等於使用者所設定之最小限制之時，此關聯法則才是被認定為有意義的 [1, 5, 6]。

### 二、Apriori 演算法

Apriori 演算法有個假設的前提，即在資料庫內的每一筆交易記錄，其所記錄的每個項目名稱必須是按照字母大小順序排列而成的[6]，例如第  $p$  筆的交易記錄中所記錄的交易項目有  $A01, A03, B02, C05$  此四項，則在資料庫中所呈現出來的即為  $t_p = \{A01, A03, B02, C05\}$ 。而整個的過程可分為兩大步驟，第一是

找出所有的高頻項目組 (Large itemset)；第二是由找到的高頻項目組推導出所有的關聯法則。

### 步驟 1. 尋找高頻項目組

```

 $L_1 = \{\text{large itemsets}\};$ 
for ( $k = 2; L_{k-1} \neq \emptyset; k++$ ) do begin
   $C_k = \text{apriori-gen}(L_{k-1});$ 
  For all transactions  $t \in D$  do begin
     $C_t = \text{subset}(C_k, t);$ 
    For all candidates  $c \in C_t$  do
       $c.\text{count} ++;$ 
  end
   $L_k = \{c \in C_k \mid c.\text{support} \geq \text{minsupp}\}$ 
End 圖 2 : Apriori_1 演算法[6,7]
Answer =  $\cup_k L_k;$ 

apriori-gen ( $L_{k-1}$ )
Insert into  $C_k$ 
Select  $p.\text{item}_1, p.\text{item}_2, \dots, p.\text{item}_{k-1}, q.\text{item}_{k-1}$ 
From  $L_{k-1} p, L_{k-1} q$ 
Where  $p.\text{item}_1 = q.\text{item}_1, p.\text{item}_2 = q.\text{item}_2, \dots,$ 
 $p.\text{item}_{k-2} = q.\text{item}_{k-2}, p.\text{item}_{k-1} < q.\text{item}_{k-1}$ 

For all itemsets  $c \in C_k$  do
  For all ( $k-1$ )-subsets  $s$  of  $c$  do
    If ( $s \notin L_{k-1}$ ) then
      Delete  $c$  from  $C_k;$ 

```

### 步驟 2. 推導出關聯法則

令  $\delta$  % 為信賴度臨界值

- 1) 對每一個在  $\cup_k L_k$  的高頻項目
- 2) Let  $X \in \cup_k L_k, Y \in \cup_k L_k, X$  是  $\cup_k L_k$  的子集合， $Y$  是  $\cup_k L_k$  其中的一個項目， $X \cap Y = \emptyset, X \cup Y = \cup_k L_k$
- 3) If ( $support(\cup_k L_k) / support(X) \geq \delta\%$ ) then 將  $X \rightarrow Y$  存入規則陣列  $R[i]$  中， $i++$
- 4) else
- 5) 將規則  $X \rightarrow Y$  丟棄
- 6) output 陣列  $R$
- 7) end

圖 3 : Apriori\_2 演算法

二、實例說明 Apriori 演算法

以下舉一簡單實例，說明 Apriori 演算法推導關聯性法則。假設資料庫  $D$  有 4 筆已按照大小排列的交易記錄，如圖 4 所示。另外，最小支持度設為 50%，最小信賴度設為 80%。

在第一次掃描的時候，找出在資料庫中所有 1-item 的項目並計算出其出現次數，收集於  $C_1$ 。而  $C_1$  中項目之出現次數超過最小支持度者（本例中次數超過 2 者），認定為高頻項目組，則收集於  $L_1$ 。

在第二次掃描的時候，先就  $L_1$  的項目組做聯集 (join) 產生 2-items 項目組，並計算其出現在資料庫的交易記錄的次數，將其記錄在  $C_2$  中，並隨後篩選出超過最小支持度的項目組，並將其記錄在  $L_2$ ；循序執行上述之步驟，直到聯集無法再產生下一批新的項目組為止。

當找完所有的高頻項目組之後，將  $L_1$ 、 $L_2$ 、 $L_3$  所有的項目組彙整，並檢視若  $(k-1)$ -項目組是為  $k$ -項目組的子集合，則將此  $(k-1)$ -項目組去除掉。例如  $\{B C\}$ 、 $\{B E\}$ 、 $\{C E\}$  是為  $\{B C E\}$  的子集合，因此  $\{B C\}$ 、 $\{B E\}$ 、 $\{C E\}$  會被刪除。最後加入支持度的因素來考慮，因此本例中所得到的關聯法則即為：

If  $\underline{B}$  and  $\underline{C}$  then  $\underline{E}$  (支持度 100%)

If  $\underline{A}$  then  $\underline{C}$  (支持度 100%)

$D$

T ID	交易資料
1	A C D
2	B C E
3	A B C E
4	B E

(a) 第一次掃描

$C_1$

項目組	次數
{A}	2
{B}	3
{C}	3
{D}	1
{E}	3

$L_1$

項目組	次數
{A}	2
{B}	3
{C}	3
{E}	3

(b) 第二次掃描

$C_2$

項目組	次數
{A B}	1
{A C}	2
{A E}	1
{B C}	2
{B E}	3
{C E}	2

$L_2$

項目組	次數
{A C}	2
{B C}	3
{B E}	3
{C E}	2

(c) 第三次掃描

$C_3$

項目組	次數
{B C E}	2

$L_3$

項目組	次數
{B C E}	2

最後所得之高頻項目組

項目組	次數
{B C E}	2
{A C}	2

圖 4：Apriori 演算法實例推導[6]

### 參、門診病歷資料問題描述

門診記錄檔所記錄的即是當每次民眾就醫求診時，醫師根據該病患所患病的症狀及需接受之治療，而在門診記錄中分別填入該次就診記錄 1 至 5 個不等的疾病項目，例如表 1 的門診資料表中，看診號為 1 的病患，就診記錄中的病碼\_1、病碼\_2、病碼\_3 分別記錄著 A34102、A31401、A34701 此三項疾病代號。藉由表 2 疾病代碼對照表加以表達，則為十二指腸潰瘍、肥厚性鼻炎及慢性肝炎此三項疾病之診斷；同樣地，看診號為 2 的病患，則被診斷出為缺鐵性貧血慢性腎功能不全及痤瘡。而所有的診斷結果中，病碼\_1 所代表的是為此次診斷之主診斷，病碼\_2 至病碼\_5 則是醫師依據病患所併發的症狀再加以記錄說明，除了是作為開立藥方之參考以外，也是為病患之所罹疾病之詳實記載。

表 1：門診資料表樣本範例

看診號	病碼 1	病碼 2	病碼 3	病碼 4	病碼 5
1	A34102	A31401	A34701		
2	A20001	A35001	A42901		
3	A101	A34103			
4	A35001	A18105			
5	A269				

表 2：疾病代碼對照表

院內代碼	國際分類碼	疾病名稱
A101	162.9	肺癌
A18105	250.9	糖尿病
A20001	280.9	缺鐵性貧血
A269	401.9	高血壓
A31401	472	肥厚性鼻炎
A34102	532.9	十二指腸潰瘍
A34103	533.9	消化性潰瘍
A34701	571.4	慢性肝炎
A35001	582.9	慢性腎功能不全
A42901	706.1	痤瘡

因此在醫生所下診斷的疾病項目之間是具有次第順序的存在，而不適宜為計算之簡便而任意更動其順序關係。另

外，由於本研究中，疾病碼項目組的組合最多僅會有 5 個項目所組成，因此會面臨到計算複雜度的問題較小。同時參考[7] 論文中對順序性資料之處理，因此就圖 2 中的 apriori-gen 其 where 條件下之  $p.item_{k-1} < q.item_{k-1}$  修改為  $p.item_{k-1} \neq q.item_{k-1}$ ，以適合本論文之資料庫特性，以下就以一個實例進一步說明。

如在圖 5 中，資料庫  $D'$  包含有 4 筆病人門診記錄在內，而每一筆記錄中的疾病代碼（如 A C D E）是為任意排列，並假設最小支持度設為 50%。

在第一次掃描，找出單個項目組出現於之頻率超過最小支持度之設定者（本例中次數超過 2 者），即被放置於  $L_1'$ 。

在第二次掃描時，先就  $L_1'$  裡所有的項目組進行聯集，產生 9 個新的項目組放於  $C_2'$  中，而 {B A}、{C A}、{C B}、{C E} 則因出現頻率超過 2，更獲選放入  $L_2'$  中。而後不斷依照此步驟進行資料庫的掃描及高頻項目組的尋找，直至找不到為止。

最後將所有高頻項目組集合起來，並剔除掉為高頻項目組之子集項目組，則本例中所得到的高頻項目組即是 {C B A} 和 {C E}。

$D'$

T ID	交易資料
1	A C D E
2	C B A
3	C B E A
4	C E

(a) 第一次掃描

$C_1'$

$L_1'$

項目組	次數
{A}	3
{B}	2
{C}	4
{D}	1
{E}	3

項目組	次數
{A}	3
{B}	2
{C}	4
{E}	3

(b) 第二次掃描

項目組	次數
{A B}	0
{A C}	0
{A E}	1
{B A}	2
{B C}	0
{B E}	1
{C A}	2
{C B}	2
{C E}	3
{E A}	1
{E B}	0
{E C}	0

項目組	次數
{B A}	2
{C A}	2
{C B}	2
{C E}	3

(c) 第三次掃描

項目組	次數
{C A B}	0
{C A E}	0
{C B A}	2
{C B E}	1
{C E A}	1
{C E B}	0

項目組	次數
{C B A}	2

最後所得之高頻項目組

項目組	次數
{C B A}	2
{C E}	3

圖 5: 修改 Apriori 演算法後之實例推導

肆、實驗結果

本論文以 816,558 筆醫院門診資料及 Apriori 演算法進行資料探勘的實驗，並將整個實驗分為以下三大步驟進行：

一、資料淨化

在門診資料表中記錄著每次病人就診時醫師所診斷出的疾病，分別記錄於病碼<sub>1</sub>、病碼<sub>2</sub>、病碼<sub>3</sub>、病碼<sub>4</sub>、病碼<sub>5</sub>的欄位中，如見表 1 所示。而其所使用的疾病代碼有門診診斷代號(A code,即以 A 為開頭的編碼法)與國際疾病分類碼 (ICD-9-CM)，因此為達到資料一致性及簡化運算的複雜度，透過院內疾病代碼對照表(如表 2 所示)，將門診資料表中的疾病代碼全部置換為國際疾病分類碼，即可得到具有一致性編碼的門診資料表，如表 3 所示。表 3 即是整個資料淨化作業所產生的資料表，同時也是資料探勘所使用的資料表。

表 3: 轉為國際疾病分類碼之門診資料表

看診號	病碼 1	病碼 2	病碼 3	病碼 4	病碼 5
1	532.9	472	571.4		
2	280.9	582.9	706.1		
3	162.9	533.9			
4	582.9	250.9			
5	401.9				

二、資料探勘

本實驗設最小支持度為 0.3% (出現的頻率為 2450)，以下表 5 - 7 則分別表示實驗所得到  $L_3$ 、 $L_2$ 、 $L_1$  高頻項目組的疾病代號及其出現的頻率。

表 5:  $L_3$

高頻項目組			出現的頻率
250.90	401.9	533.90	3407
460	487	490	3778
715.90	729.2	733.00	2494
715.90	733.00	533.90	2503
722.10	715.90	729.2	2557

表 6 :  $L_2$ 

高頻項目組	出現的頻率
250.90	272.0
250.90	401.9
250.90	414.9
250.90	465.9
250.90	533.90
250.90	564.0
250.90	571.40
250.90	780.52
401.9	250.90
401.9	414.9
401.9	465.9
401.9	533.90
401.9	564.0
402.90	533.90
460	487
460	490
461	460
463	564.9
486	460
486	487
486	490
487	490
715.90	722.10
:	:
:	:
729.2	733.00
733.00	533.90

表 7 :  $L_1$ 

高頻項目組	出現的頻率
112.9	4873
242.9	3212
250.6	6537
250.9	49469
272	9312
274	5510
274.9	4747
285.9	2797
300	4291
306.9	3139
:	:
:	:
884	5639
894	6182
V20.2	14073
V22.1	27489
V78.2	3427

並設定信賴度為 90%，因此所得到的關聯法則如表 8 所示。

表 8 : 關聯法則表

關聯法則	信賴度%
460 and 490 $\Rightarrow$ 487	92.83
729.2 and 733.00 $\Rightarrow$ 715.9	90.00

### 三、探勘資料之釋義

將表 8 所得到的關聯法則，藉由表 9 的代碼轉換表將疾病之代碼與名稱做對照，即可產生下列 2 則人類所能理解的知識法則：

法則 1: If 急性鼻咽炎 and 支氣管炎 then 流行性感冒

由於流行性感冒通常都會伴有急性鼻咽炎及支氣管炎的症狀發生，因此病患若有急性鼻咽炎與支氣管炎的症狀出現時，我們可推斷此人很可能是罹患了流行性感冒。

法則 2： If 神經痛、神經炎及神經根  
and 骨質疏鬆症 then  
骨關節病

而當病患有有神經痛、神經炎及神經根症狀出現時，而本身又具有骨質疏鬆症存在的話，則我們將可以推斷此病人很可能得到骨關節病。

經過與醫院醫師討論後，以上的兩條關聯法則可成立，證明此研究成果值得院方當參考。

表 9：疾病代碼與疾病名稱對照表

疾病代碼	疾病名稱
460	急性鼻咽炎
461	急性鼻竇炎
487	流行性感冒
490	支氣管炎
715.9	骨關節病
729.2	神經痛、神經炎及神經根
733.00	骨質疏鬆症

## 伍、結論及後續研究

本論文利用若瑟醫院的門診資料庫，並運用修改過之 Apriori 演算法進行門診疾病之關聯法則研究。在醫院門診資料的知識發現上，我們得到共 2 則有用的知識，此 2 則知識法則，不僅通過支持度 0.3% 本研究所設定的門檻值，並且具有 90% 以上的可信度，同時也經醫師驗證是為合理可行的。這些知識所顯示疾病之間的關聯性，不僅可以供醫師診斷時的參考，亦能提供一般民眾在自我保健的疾病防治方面有預警功效，甚至能提供關於醫務管理的決策資訊，讓醫院能提供更好的醫療品質造福大眾。因此，未來的研究將再努力的目標有以下幾點：

1. 進行更大量資料之資料探勘工作，以找出更有效與更豐富的知識法則。
2. 將時間的因素加入考慮，藉由追蹤病人的病史，以找出在長時間的累積下，疾病之間的關聯性，即尋找出循序樣型的關聯性法則。
3. 找出疾病診別之間的關聯性，以作為

醫務管理決策參考。

4. 研究新的演算法，以減少計算複雜度的問題。
5. 就病患之背景分析，以找出容易引發某疾病之族群特徵。

## 致謝

感謝雲林縣虎尾鎮聖若瑟醫院提供實驗所需之門診資料。

## 六、參考文獻

1. 王仁傑, "效找出較常資料項目型樣的關聯法則之研究," 逢甲大學資訊工程所碩士論文, 民國 89 年 6 月.
2. W. J. Frawley, G. Piatetsky-Shapiro, and C. J. Matheus, "Knowledge discovery in database: an overview," *Knowledge Discovery in Database*, AAAI/MIT Press, pp. 1-27, 1991.
3. G. Piatetsky-Shapiro (Editor), *Knowledge Discovery in Database*, AAAI/MIT Press, 1991.
4. R. Agrawal, T. Imielinski, and A. Swami, "Database mining: a performance perspective," *IEEE Trans. Knowledge and Data engineering*, vol. 5, no. 6, pp. 914-925, Dec. 1993.
5. R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large database," *The 1993 ACM SIGMOD Conference*, pp. 207-216, 1993.
6. R. Agrawal, and R. Srikant, "Fast algorithm for mining association rules," *Proc. of the 20th VLDB Conference*, pp.487-499, 1994.
7. R. Agrawal and R. Srikant, "Mining Sequential Patterns," *Proc. of the Int'l Conference on Data Engineering*, Mar. 1995.
8. G. J. Klir and B. Yuan, *Fuzzy Set and Fuzzy Logic: Theory and Applications*, Prentice Hall, 1995.
9. P. Adriaans and D. Zantinge, *Data*



- Mining*, Addison Wesley, 1996.
10. U. M. Fayyad, G. Piatetsky-Shapiro, P.Smyth, and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, CA, 1996.
  11. M.J.A. Berry, and G. Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Support*, Wiley Computer Publishing, 1997.
  12. U. Fayyad, "Mining database : towards algorithms for knowledge discovery," *IEEE Computer Society Technical Committee on Data Engineering*, vol. 21, no.1, pp. 39-48, 1998.
  13. T.-P. Hong, C.-S. Kuo, and S.-C. Chi, "Mining association rules from quantitative data," *Intelligent Data Analysis* 3, pp. 363-376, 1999.
  14. P. C. Pendharkar, J. A. Rodger, G. J. Yaverbaum, N. Herman, and M. Benner, "Association, statistical, mathematical and neural approaches for mining breast cancer patterns," *Expert Systems with Applications*, vol. 17, pp. 223-232, 1999.
  15. B. Liu, W. Hsu, and Y. Ma, "Pruning and summarizing the discovered associations," *Proc. of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 125-134, Aug. 1999.
  16. S. Parthasarathy, M. J. Zaki, M. Ogihara, and S. Dwarkadas, "Incremental and interactive sequence mining," *Proc. of the eighth international conference on Information knowledge management*, pp. 251-258, Nov. 1999.
  17. R. Elmasri and S. B. Navathe, *Fundamentals of Database Systems*, third edition, Addison-wesley, 2000.
  18. 若瑟醫院全球資訊網 : <http://www.stjoho.org.tw/>
  19. 行政院衛生署八十九年度醫院評鑑暨教學醫院評鑑合格名單  
<http://www.doh.gov.tw/new/news/90/900111-1.htm>