

# 關聯法則之互動增強型挖掘模式的設計與實作

## An Intension Mining Model for Mining Association Rules

邱宏彬<sup>1</sup>

楊昇樺<sup>2</sup>

<sup>1</sup>南華大學資訊管理系助理教授

<sup>2</sup>南華大學資訊管理研究所碩士

### 摘 要

傳統知識發掘過程常針對單一問題來討論與解決，但現實上，整個知識發掘是一個互動繁覆，而且必須整合人工處理程序及機器處理程序的過程；在這整個過程中所遇到的問題，必須一併來討論及解決才有意義，在本論文中，我們稱之為互動增強型知識發掘模式(Intension Mining Model for KDP)。我們針對關聯法則挖掘，設計出一個基於互動增強的知識發掘模式之整合式構架；它能針對企業或研究需要，進行一個或多個問題的解決，包括：漸進式挖掘、線上挖掘、使用者有趣性挖掘等，使得關聯法則挖掘更具資料合用性與實用性。

關鍵詞：互動增強型知識發掘模式、資料挖掘、關聯法則、漸進式挖掘



### Abstract

The traditional data mining approaches often aim at discussing and addressing a sole question. However, researchers and practitioners have realized that KDP is an interactive-intensive and semiautomatic process, which is so called Intension Mining Model for KDP. A number of algorithms have been proposed for association rule mining, including incremental mining, online mining, interestingness mining, and so on. In this paper, we explore the extended model for association rule mining that is one of the popular research areas in data mining. An interactive and integrated framework is designed and implemented, and several experiments are conducted to verify the effectiveness and usefulness.

**Keywords:** Intension Mining Model for KDP, data mining, association rules, incremental mining

### 壹、緒論

將有用的知識或是法則，從資料庫系統中挖掘出來的整體過程，就是知識發掘的流程 (Knowledge Discovery Process)(Chen et al., 1996)。真正的知識發掘分為(1)資料的取得與淨化階段、(2)資料倉儲階段、(3)知識擷取階段、(4)規則詮釋階段。資料挖掘的結果必須經過此一領域相關的學者加以詮釋，並確認資訊的可用性，以便將新奇有用且讓使用者感興趣的資訊取出，並用一般大眾都能了解的名詞解釋出來，此一階段就是規則詮釋的階段。經過了一般化的資訊或是規則，就是一般所說的知識或是規則。因為資料挖掘是整個過程中最重要的核心，所以也有人將知識發現過程稱之為資料挖掘流程，如圖 1 所示。

近年來，許許多多的研究都把重點

放在資料挖掘演算法效能的改進，強調機器處理程序的部份，但往往卻忽略掉了人工處理及人為介入部份的重要 (Chen et al., 1996; Goebel et al., 1999)。能快速把使用者所想要的特徵項目 (Pattern) 利用網路或媒介提供給使用者才是最後的目的。在 (Gupta et al., 2000) 中，有學者分析實際上知識發掘的流程，將傳統的挖掘模式加上人力介入的部份及修改資料處理部份，使之更符合現實狀況，稱之為互動增強型知識發掘流程模式 (Intension Mining Model for KDP, IMM)，如圖 2 所示。各個處理流程 (Process) 說明如下：

IM1：由某種動機引發專業人員來進行執行整個知識發掘的過程，經過專業領域人員設計、定義整個挖掘流程，設立各個階段的目標，並針對結果評估是否再挖掘。



IM2：由 IM2a 至 IM2c 構成，它與傳統流程 P1 至 P3 是一樣的。真實上資料庫會隨著時間遞增，所以這步驟應該會週期性地重覆 IM2。

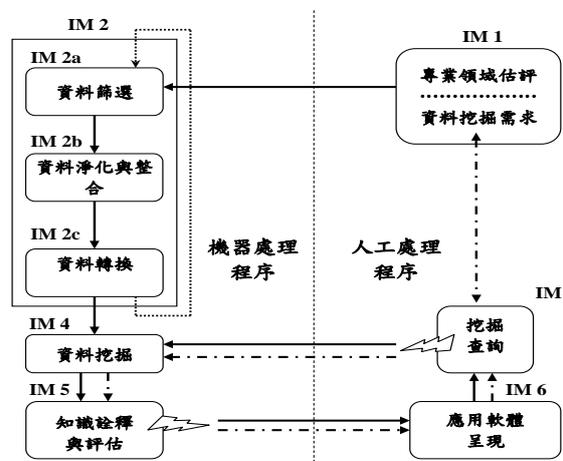
IM3：使用者所下的查詢動作。這個程序是 IM4 運作的一個依據，但是我們無法預先得知何那一種類的使用者會何時利用何種媒介來下何種範圍的查詢動作。

IM4：這部份便是使用到資料挖掘演算法來實地進行資料挖掘的程序。這是最重要的核心部份。

IM5：知識產生的程序，與傳統流程 P5 是一樣的。

IM6：透過某種方式呈現，讓使用者清楚了解到所挖掘到的法則，讓使用者自行評估這些知識適不適用。

IMM 提供使用者對挖掘正確的結果具有聚焦(focus)的能力，而這種能力建立在使用者現有的知識基礎上。因此本研究提出基於 IMM 之整合性架構，讓使用者可以將其現有的知識加入挖掘的過程中，去協助發掘所需的



- 線表示資料流程。
- ..... 線表示周期性重覆的流程。
- - - 線表示可能會重覆的流程。

圖 2、互動增強型知識發掘流程模式

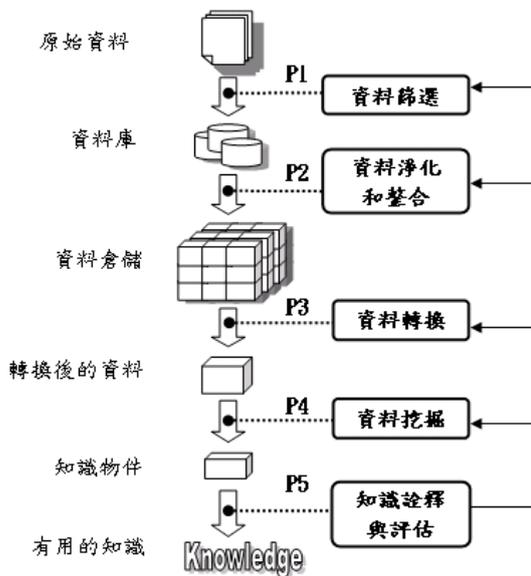


圖 1、傳統的知识發掘流程模式

知識。資料挖掘的工作有許多種，如分類、預測、分群、關聯法則等 (Goebel and Gruenwald, 1999)。想要設計一個滿足所有 Data mining tasks 的 IMM 是一件非常困難的工作。目前 association rule mining 是學術界及企業間熱門的領域，因此，本文本研究將只針對 association rule



mining，嚐試提出一個植基於 IMM 的整合性架構，以輔助使用者快速有效的挖掘出有用的關聯法則。

## 貳、關聯法則挖掘演算法之相關研究

### 2.1 Apriori-like 關聯法則挖掘

Apriori 演算法首先於 1994 年由 Agrawal et al. 提出，包含了兩個重要的步驟：(1)反覆的產生候選項目組和搜尋整個資料庫，直到找出所有的大項目組。(2)利用(1)所找出的大項目組，推導出所有的相關法則。因為 Apriori 演算法牽涉到多次資料庫掃描以及可能產生過多的候選項目集，因此是一件耗時的工作。有許多改善的方法被提出，如 DHP 演算法 (Park et al., 1995) 利用減少候選目集的個數來減少計算量，以增進效能，而 DIC (Brin et al., 1997)、Sampling (Toivonen, 1996) 等演算法則是減少資料庫的掃描次數來改善效率。在 Sampling 方法中，樣本數及資料分佈的範圍均會影響資料挖掘的結果，相對的其可信度也會有所偏差。但 Toivonen 認為在可接受的情形下，犧牲一些可信度來換取好的處理效率是非常值得的。

### 2.2 漸進式關聯法則挖掘

Apriori 演算法不適合漸進式探勘，因為每次新的交易資料進入資料庫時，這些演算法必須重新掃描更動後的整個

資料庫，而不能只考慮新增記錄集。Cheung et al. 所提出的 FUP (Fast Update) 演算法 (Cheung et al., 1997) 可算是漸進式探勘方法的代表範例，當資料庫更動時，藉由比較原始資料庫的高頻項目組 (Large itemsets) 與新增記錄集中的高頻項目組來減少掃描原始資料庫的次數。

### 2.3 線上關聯法則挖掘

在許多研究利用縮減項目集絡 (Reduced Itemsets Lattice) 的觀念來解決線上關聯法則探勘的問題。該方法利用 Apriori principle 將必要的項目集資訊儲存在絡 (Lattice) 的架構中，以方便關聯法則探勘，若儲存的資訊越詳細，探勘時就能減少必須到資料庫做確認的情況。

### 2.4 稀少項目問題

以一個賣場中的家電和零食來看，家電因為價錢較高且需求量較低，當門檻值定較高時，家電的相關法則很容易被犧牲掉。Liu et al. 在 1999 便提出一個能解決稀少項目問題的方法，稱之為 MSApriori 演算法 (Liu et al., 1999)。這個演算法能讓使用者設定每一個項目的最小支持度 (MIS, Minimum Item Support)，因此每一個項目都有自己的最小支持度，當單一項目要結合成更高候選項目集的時候，依各項目選取最小的支持度來作為自身的最小支持度。



## 2.5 使用者有趣性問題

在進行關聯法則挖掘時就像一個黑箱，使用者能干涉到挖掘的能力很低，唯有等待結果出來，使用者如果覺得結果不好，也只能重新再一次挖掘；因此讓使用者對於自己所想要的條件來多重限制挖掘的執行範圍才能解決此一問題。

Yen et al. 在 1997 便提出 EDM (Efficient Data Mining) 演算法。簡單來說 EDM 演算法提出一個資料挖掘語言，舉例如圖 3。它能幫助使用者來限制搜尋規則的範圍，使用者可以選擇要何種項目的關聯法則，也就是可自行選擇從那一資料表、何種項目為前項和後項、最小支持度為多少。如此即可以確定找出來的法則是使用者真正所需求的法則。

```

Mining Association Rules
From <Database>
With
  (Antecedent <Items> (*))
  (Consequent <Items> (*))
Support s
  
```

圖 3、EDM 的需求語法

## 2.6 使用者互動性挖掘查詢

一般來說，當使用者想要進行查詢時，絕大部份會執行一連串的挖掘查詢 (Mining Query) 來進行挖掘，一直到使用

者的需要被滿足為止。也因此整個挖掘過程中，使用者所下的一連串查詢，通常都是很相似的；Czejdo et al. 發現到這樣子的一個情況，便利用實體化視域 (Materialized View) 企圖來解決這樣一個問題。

## 參、問題描述與系統目標

本研究擬發展一套關聯法則挖掘查詢演算法，可以處理使用者互動性查詢的問題，本研究之具體目標有：

(1) 企圖解決線上挖掘、稀少項目和使用者有趣性的問題，要解決這些問題，必須先將必要資訊存於資料倉儲中，以供隨時搜尋之用，因此，本研究設計出關聯法則之多層更新挖掘法 (Multi-layer Update Miner, MUM)。

(2) 為能利用 MUM 存在資料倉儲中的資訊，快速回應使用者的互動性查詢，而且能提供使用者可選擇系統自動設定動態最小支持度參數或自行針對各層設定支持度，因此本研究以現有關聯法則中的 Apriori 演算法為基礎，再加上參考動態最小支持度演算法及 Sampling 演算法的觀念來設計出關聯法則查詢演算法 (Query-based Association Rule Miner, QARM)。

(3) 設計出一套能利用 MUM 和 QARM 來有效地挖掘出具有實用價值也就是使用者真正需要之法則。本研究稱



之為適性化查詢系統 (Query-based Adaptive MUM, QAMUM)。

#### 肆、系統架構與功能

本研究提出一套完整的系統 QAMUM，系統架構如圖 4 所示。

##### 4.1 使用者溝通介面

在挖掘過程，需要使用者親自設定參數與限定資料範圍，以達到最有效的挖掘結果。因此，建立良好使用者溝通介面應

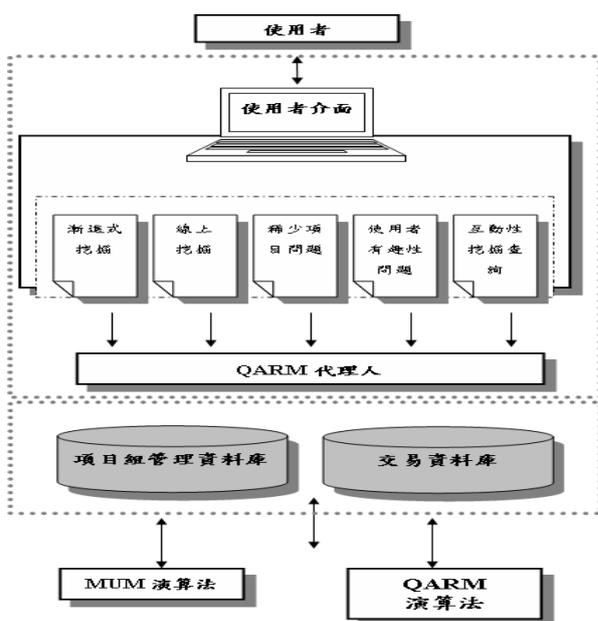


圖 4、QAMUM 系統架構

著重在使用者的溝通模式與使用者可能遇到之問題與種類。

##### 4.2 QARM 代理人

QARM 執行的工作主要為判斷使用者所下的查詢是屬於那一類問題，但使用者的查詢可能結合多個問題，因此每個問題處理程序雖各自獨立，但可相互支援處理。各處理程序如下：

###### (1) 漸進式挖掘問題:

呼叫 MUM 演算法，將新增資料庫交易部份進行處理，處理完結果存進資料倉儲中及回傳給使用者。

###### (2) 線上挖掘問題:

直接搜尋資料倉儲，將滿足門檻值的法則回傳給使用者。

###### (3) 稀少項目問題:

判斷門檻值參數，看使用者希望系統產生動態門檻值或自行多層設定。

###### (4) 使用者有趣性問題:

依使用所下的條件，直接搜尋資料倉儲，將符合的資訊回傳給使用者。

###### (5) 互動性查詢問題

此問題牽涉到的範圍較大，有可能會跨越到其他問題層面，它會有四種發生情況，我們針對各個情況作處理：

###### (一) 查詢條件與原資料庫相符

依使用所下的條件，直接搜尋資料倉儲，將符合的資訊回傳給使用者

###### (二) 查詢條件與部份原資料庫相符

將所要用到的資訊從資料倉儲中過濾出來。呼叫 QARM 演算法進行處理，將結果回傳給使用者，並將結果存在實體化視域，留待下一次查詢使用。



(三) 查詢條件與新增交易資料相符  
呼叫漸進式挖掘問題處理程序。

(四) 查詢條件與部份原資料庫和新增交易資料相符

將情況(二)和情況(三)合併處理，並將結果相加，將符合門檻值的法則回傳給使用者。

#### 4.1 MUM 演算法

本研究利用先前提出的 MUM 演算法(楊昇樺, 2002)作為本系統之資料挖掘方法的基礎。因為它較有彈性，能解決多種的問題，如：線上挖掘、動態挖掘。再將挖掘完的相關資訊存回資料倉儲中，提供企業作更多潛在資訊的分析，而且查詢發生時，也能再一次利用到這些資訊，加快挖掘的速度。

#### 4.1 QARM 演算法

本研究把 QARM 演算法建構在 MUM 演算法之上，來解決快速回應使用者查詢的問題。利用 MUM 演算法執行完後所存在資料倉儲裡的特徵項目，來產生預測大項目集，解決 Apriori-Like 必須產生候選項目集，需要多次掃描資料庫來挖掘法則的主要問題。

由於挖掘時間有限，因此本研究利用系統已處理完整個資料庫後的項目組資訊，來解決使用者互動性查詢的問題。在統計學觀念中，是以全體研究對象為母體(Population)，透過抽樣抽出的

部分研究對象為樣本(Sample)；因對於母體限於有限的人力、物力及時間的限制下，無法對全體的研究對象進行調查，故利用樣本來推估出母體。但在此研究中，因為我們不要使用者每下一次查詢(抽樣)，就重新挖掘一次，因此我們利用 MUM 處理完後的資訊，來估計查詢資料的項目組(樣本)的支持度，母體、樣本彼此間的關係圖如圖 5。

本研究使用了預測大項目集的做法，因此可加快挖掘速度。也因為如此使得我們推導出來的法則會有小部份遺失的可能。另一方面使用 Apriori-Like 的演算法方式也使得本研究應用度增廣，我們可以針對門檻值做多種應用設定，而不需要改變其挖掘方法。

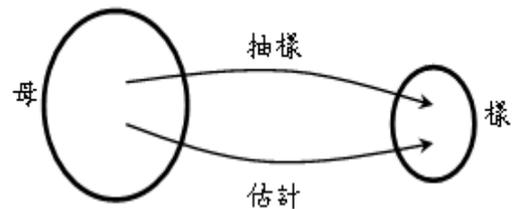


圖 5、母體、樣本關係圖

QARM 處理的過程如下：

- (1)先確定過濾規則強度的門檻值—動態門檻值或使用者自訂的各層門檻值。
- (2)先掃描一次資料庫，根據門檻值找出樣本的 1-高頻項目組。



- (3)判斷樣本的 k-高頻項目組與母體的 k-高頻項目組是否一致(k=1)，一致的部份則直接利用 MUM 資料表的資訊。
- (4)並利用樣本高頻項目組結合(Join)產生 2-預估候選項目組，再根據門檻值找出 2-預估大項目組。
- (5)  $k = k + 1$ ，重覆執行步驟 3；直到不再結合出新的預估候選項目組為止。
- (6)將產生的預估大項目組，再掃描一次部份資料庫，找出其真正 count 次數，並判斷是否滿足門檻值。
- (7)將所有的大項目組，根據最小信賴度推導出其相關法則。

當對樣本作推論時，本研究先掃描一次樣本資料庫找出 1-高頻項目組的支持與母體作一比對時，會有兩種情形的發生，我們便針對這兩種情形分開來探討。

**情況一：**樣本與母體中的 1-高頻項目組的支持度有相同的情形。

假設單一項目 A 有此情況發生時，唯一的情況便是母體中的有項目 A 發生的交易資料集全被我們選取中了，因此母體與樣本中全部有關 A 的超集合(SuperSet)皆一致。

**情況二：**樣本與母體中的 1-高頻項目組的支持度有不相同的情形。

在此情況下，本研究希望能利用母體的資訊來推論出樣本的長度 2 以上的項目集的支持度。因此最好的情形下，就是能非常接近實際的支持度但不要低於它，有低於實際支持度的情形時，就有可能會被刪除。因此，我們提出了預測項目組支持度(Predictive Itemset Support, PIS)公式。

本研究假設母體資料庫 P 包含了交易資料集 T 和全部項目的集合 I，我們選取了交易資料集 X 作為樣本資料庫 S。A 是單一項目或是連續項目集  $A_i \wedge \dots \wedge A_j$ ，當  $A_i, \dots, A_j \in I$ 。 $\sigma_X(A)$  是 A 在交易資料集 X 的支持度，當  $X \in T$ ；為了盡量能不產生比實際支持度還低的預估值，因此我們皆選取較大的相關樣本單一項目支持度來乘上較大可能發生的機率成為可能項目集支持度(Possible Itemset Count, PIC)其公式如下：

$$PIC(A) = \left[ \text{Max}(\sigma_X(A_i), \dots, \sigma_X(A_j)) * \text{Max}\left(\frac{\sigma_T(A)}{\sigma_T(A_i)}, \dots, \left(\frac{\sigma_T(A)}{\sigma_T(A_j)}\right)\right) \right] \quad (1)$$

但在上述的同時，有可能估計出來的值過大，喪失了我們要推估到逼近真實支持度的精神，因此本研究把公式擴大成：



$$PIS(A) = \min \{ \sigma_X(A_1), \dots, \sigma_X(A_j), PIC(A) \}$$

(2)

我們再進一步探討門檻值、實際支持度和估計支持度之間的關係以及對於本研究會有何影響，如表 1。

表 1、門檻值與支持度的關係與影響

發生狀況		相互關係		處理動作
		RS	PIS	
1	$PIS > RS > MS$	高頻	高頻	正確不作處理
2	$PIS > MS > RS$	非高頻	高頻	高估，須再掃描一次資料庫，將此誤判非高頻的項目組過濾掉
3	$MS > PIS > RS$	非高頻	非高頻	正確不作處理
4	$MS > RS > PIS$	非高頻	非高頻	正確不作處理
5	$RS > PIS > MS$	高頻	高頻	正確不作處理
6	$RS > MS > PIS$	高頻	非高頻	低估，可由掃描資料庫，將刪除(遺失)的高頻項目組補回

門檻值：MS，實際支持度：RS，估計支持度：PIS

狀況 1、5、3、4，不會造成任何的影響，所以不予處理。在狀況 2 下，必須再掃描一次資料庫，把估計錯誤的項目組刪除。在狀況 6 下，因原本項目組應為高頻，估計太低成為非高頻而刪除掉，這種情況雖然很少，但有發生的可能性，因此我們提出一個衡量參數  $\alpha$ ，當有發生狀況 6 的情形，必須滿足公式才可允予刪除。

$$\text{衡量公式} = \frac{|MS - PIS|}{MS} > \alpha \quad (3)$$

QARM 有兩個主要的步驟如圖 6 所示。

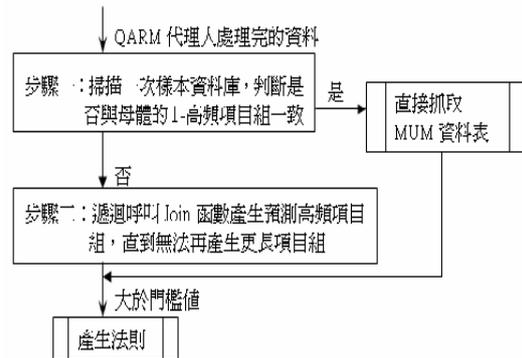


圖 6、QARM 演算法

## 伍、實驗結果分析

假設某一資料庫編號為 T10I100D10k，代表此資料庫包含 10000 筆交易，資料庫中所包含單一項目個數為 100，每筆交易平均包含 10 個項目。本研究用來測試的資料庫為 T10I50D10k、T10I100D10k、T10I100D5k。實務上，門檻值的決定依不同的問題而有所差異。本研究利用不同的門檻值，來實驗和探討本系統的執行效能。

### 5.1 Apriori、MUM 和 QARM 的效率比較

本研究將 Apriori、MUM 和 QARM 來處理 T10I50D10k、T10I100D10k 兩個



資料庫，其執行效率的結果如表 2 所示。

表 2、各演算法執行效率時間表(秒)

T10I50D10k			
預設門檻值	0.20%	0.30%	0.40%
Apriori	1303	1186	1178
MUM	2875	2875	2875
QARM	632	341	342
T10I100D10k			
預設門檻值	0.20%	0.30%	0.40%
Apriori	6824	1523	392
MUM	3384	3384	3384
QARM	925	473	181

由表 2 之數據，可以知道不同的門檻值對於 MUM、QARM 和 Apriori 的影響：

- (1) MUM 演算法不管門檻值多少，均須把所有交易筆數拆解完存到資料表中，因此處理速度會被拖慢。
- (2) 可看到變化最大為 Apriori，因在資料庫中所存在的項目愈少，能產生的候選項目組就愈少，也就減少更多掃描資料庫的機會。當門檻值高到一定程度時，Apriori 演算法會比 QARM 演算法處理時間上更為快速。
- (3) 可觀察出 QARM 演算法的處理效率最好，尤其在門檻值愈低時，愈能看出它減少掃描資料庫所出現的效果，而它主要所花的時間集中在掃描 MUM 主資料表中。

## 5.2 動態新增資料庫之漸進式挖掘比較：

Apriori 演算法採批次處理的方式運作，所以假設原始資料庫已有 10000 筆交易記錄，之後依序新增 1000、500、100 筆。門檻值設定為 0.3%。其執行時間如下表 3。

表 3、漸進式挖掘時間比較表(秒)

交易筆數	10000筆	11000筆	11500筆	11600筆
Apriori	1523	1308	1307	1288
MUM	3384	289	138	24

在第一次挖掘後，由於 MUM 僅需處理新增的資料，從表中可明顯的看出其挖掘時間不會隨著資料量的增加而遞增。反觀 Apriori 演算法，由於每次新增資料時均需重新挖掘異動後的整個資料庫，因此挖掘時間只會愈來愈增加。事實上，在現實生活中交易資料是漸增的，所以在交易的同時 MUM 即可同步進行項目組資料表的更新，無須一次處理大筆資料，這也是 MUM 的特性。

## 5.3 線上探勘、使用者有趣性問題及稀少項目問題：

5.3 以上這些問題其解決方式皆只需要搜尋項目組資料表，本研究以假設使用者自行動態設定門檻值為例。在資料庫的交易記錄為 10000 筆時，門檻值設定分別為 0.2%、0.3%、0.4%。其結果如表 4。



表 4、調整門檻值時間比較表(秒)

調整門檻值	0.20%	0.30%	0.40%
Apriori	6824	1523	392
MUM	2	2	2

由於每次給定新的門檻值時 Apriori 演算法需要重新掃描資料庫以挖掘所有高於支持度的項目組，可以明顯的看出其挖掘時間隨著門檻值的下降而快速遞增。但 MUM 隨時保有最新的項目組資訊，所以當門檻值改變時，僅需查詢項目組資料表，無須重新挖掘整體資料庫，即可找出所有超過門檻值的大項目組。因此，其挖掘時間不會隨著門檻值變化。由實驗數據可以明顯的看出 MUM 的優勢。

#### 5.4 QARM 高頻項目組遺失的探討

本研究採用 T10I100D5k 資料庫來進一步探討 QARM 發生高頻項目組遺失的情況。結果如表 5、6。

表 5、遺失率與 參數關係表

門檻值	遺失率 $\alpha=0$	$\alpha=\frac{1}{15}$	$\alpha=\frac{2}{15}$	$\alpha=\frac{3}{15}$	$\alpha=\frac{4}{15}$	$\alpha=\frac{5}{15}$
0.2	1.02%	0.32%	0.18%	0.14%	0	0
0.3	10.08%	3.79%	1.29%	0.38%	0	0
0.4	29.47%	17.83%	8.33%	2.92%	0.58%	0.14%

表 6、執行效率時間比較表

門檻值	Apriori	QARM
0.2	3257	436
0.3	1808	254
0.4	238	116

由表 5 中，通常設到  $\alpha=4/15$  時，已接近零遺失率了，由表中可得知當門檻值愈高的話，遺失率愈高，我們的參數必須設得愈高，才能降低遺失率，因此在執行速度和遺失率發生的情況上，要做一抉擇。經由實驗結果，我們會建議在實務上，當門檻值偏高或抽樣樣本過小時，應選擇 Apriori；如此，在處理效率上與 QARM 相差不多，且不會有高頻項目組遺失的情況出現，如表 6 所示。

#### 陸、結論與未來展望

本文提出一個資料挖掘系統的架構—QAMUM 並加以實作驗證，此架構能配合使用者需要解決多個知識發掘流程上的問題，大大加大了其實用性。這些問題包含：漸進式挖掘問題、線上挖掘問題、稀少項目問題、使用者有趣性問題、互動性查詢問題。實驗結果顯示 QAMUM 在上述所提的各個問題上均能有效的解決。利用平常所累積的關聯法則資訊，能有效地解決使用者查詢問題，這是一個新的想法與實作的研究，但計算時間與儲存空間的取捨問題及其



可能的改善方式和如何將所儲存的資訊作更有用、有效的應用是以後進一步研究的重要課題。

### 參考文獻

- [1] 楊昇樺、毛立仁、邱宏彬，“關聯法則之多層更新挖掘法及其應用”，第三屆產業資訊管理學術暨新興科技實務研討會論文集，pp.549~557，2002。
- [2] B. Czejdo, M. Morzy, M. Wojciechowski and M. Zakrzewicz, “Materialized views in data mining,” Proc. of IEEE the 13th International Workshop on Database and Expert Systems Application, pp.827, 2002.
- [3] B. Liu, W. Hsu, and Y. Ma, “Mining association rules with multiple minimum supports,” Proc. of the fifth ACM-SIGKDD Conf. on Knowledge Discovery and Data Mining ACM, pp.337-340, San Diego, CA, 1999.
- [4] D.W. Cheung, J. Han, V. Ng, and C.Y. Wang, “A general incremental technique for maintaining discovered association rules,” Proc. of Conf. Sys., pp.185-194, 1997.
- [5] H. Toivonen, “Sampling large databases for association rules,” Proc. of the 22nd Conf. on VLDB, pp.134-145, 1996.
- [6] M.S Chen, J. Han, and P.S. Yu, “Data Mining: An overview from a database perspective,” IEEE Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, pp.866-883, December 1996.
- [7] M. Goebel, and L. Gruenwald, “A survey of data mining and knowledge discovery software tools,” ACM-SIGKDD, Val. 1, pp.20- 33, June 1999.
- [8] Pauray S.M, and Chien-Ming Chen, “Mining interesting association rules from customer database and transaction database, “ Information Systems, 29(8), pp.685-696, 2004
- [9] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” Proc. of the 20th Conf. on Very Large Databases, pp.487-499, Santiago, Chile, 1994.
- [10] S. Brin, R. Motwani, J.D. Ullman, and S. Tsur, “Dynamic itemset counting and implication rules for marketing basket data,” ACM-SIGMOD Conf. on Management of Data, pp.255-264, 1997.
- [11] S.J. Yen and L.P. Chen, “An efficient data mining technique for discovering



- interesting association rules,” Proc. of the 8th International Workshop on Database and Expert Systems Application, pp.664-669, 1997.
- [12] S.K. Gupta, V. Bhatnagar, S.K. Wasan and D. Somayajulu, “Intension mining: a new paradigm in knowledge discovery,” Technical Report No. IITD/CSE/TR2000/001, March 2000.
- [13] Y.-J. Tsai, J.-Y. Chiang, CBAR: an efficient method for mining association rules, Knowledge-Based Systems, pp. 99-105, 2005.

