



經濟部標準檢驗局 98 年度自行研究計畫

進口貨品抽批機制風險評估

研究人員姓名：標準檢驗局臺中分局技士 朱延朗
標準檢驗局臺中分局技士 林東昇
嶺東科技大學助理教授 張翠蘋

經濟部標準檢驗局台中分局 編印
中華民國 99 年 01 月 05 日

標準檢驗局臺中分局 98 年度自行研究 報告提要表		填表人：朱延朗、林東昇、張翠 蘋 填表日期：99.01.06	
研究報告名稱	進口貨品抽批機制風險評估		
研究單位 及研究人員	標準檢驗局台中分局 嶺東科技大學	研究 期程	自 98 年 01 月 01 日 至 98 年 12 月 31 日
報 告 內 容 提 要			
<p>1. 現行的進口貨品抽批機制對於任何廠商所進行的抽批檢驗作業是採取混合抽批，此種作法容易造成現有抽批及檢驗人力無法有效的發揮，合格率高進口商不斷被抽驗，而合格率低進口商卻屢次避過抽驗，在實務上也曾發生多次貨主抱怨為何他的貨物總是一直被抽中。如此的抽批機制下造成進口貨品的控管品質低落，參考現行海關對於進口績優廠商享受較低的抽驗比率，而將人力放在控管不良廠商身上。因此，本研究的動機在於以有限的人力成本下，考量進口商過去所累積抽驗貨品的合格率，提昇進口貨品的控管品質。</p> <p>2. 目的：在不增加現有人力成本條件下，考量進口商過去所累積抽驗貨品的合格率，提昇進口貨品的控管品質。採用資料挖掘(Data Mining)的技術，在現有的進口商報驗資料庫中挖掘出進口廠商、進口貨品與報驗貨品合格率之間關係。</p> <p>3. 方法：</p> <p>(1). 在現有的進口商報驗資料庫中，搜集累積大量進口商在過去進口貨品的項目及合格與否等資訊。</p> <p>(2). 在現有的資料挖掘技術，以關聯式規則(Association Rules, AR)為核心，分析各種不同資料挖掘演算法是否適用於挖掘進口商報驗資料庫。</p> <p>(3). 透過分析，決定採用的資料挖掘演算法，並以此演算法進行測試挖掘，驗證是否可以找出進口商、進口貨品與報驗貨品合格率之間關係。</p> <p>(4). 設計一風險評估機制，將透過資料挖掘所產生的進口商、進口貨品</p>			

與報驗貨品合格率之間關係加入進口商抽批檢驗作業中。

(5). 反覆測試此一機制，檢討是否合用。

4. 發現與建議：在我們的實驗中，我們發現 96 年及 97 年中部地區裕毛屋是目前最大的進口廠商，此外，我們亦可判斷，夏暉物流有限公司是家有品質的進口商，其進口貨品的合格率很高。對於一些進口商例如奇銳、台灣大等經常進口不合格產品，本局應將人力控管在這些進口商所進口的產品上。

目錄	
摘要	1
1. 前言	1
2. 相關研究	3
2.1 資料探勘關聯法則原理	3
2.2 標檢局之商品報驗程序及資料	5
3. 研究步驟及方法	6
4. 實證分析	7
4.1 資料前處理	8
4.2 以關聯法則探勘標檢局之商食品檢驗交易資料庫	8
4.3 實驗結果	12
4.3.1 基本工具及參數說明	12
4.3.2 實際商品檢驗資料庫結果分析	12
4.3.2.1 以關聯法則探勘的流程	13
4.3.2.2 96 年廠商報驗記錄探勘結果與分析	16
4.3.2.3 97 年廠商報驗記錄探勘結果與分析	16
4.3.2.4 96 年及 97 年廠商報驗記錄探勘結果與分析	17
5. 結論與未來研究方向	18
6. 參考文獻	19

摘要

本研究是以資料探勘關聯法則技術針對經濟部標準檢驗局的進口廠商報驗貨品資料庫進行分析與預測。我們透過現有資料探勘技術中的關聯法則演算法對進口廠商所進口的貨品進行資料挖掘並找出廠商、進口貨品、生產地及其送驗合格率之關聯規則。透過資料挖掘關聯法則技術，我們篩選出送驗貨品合格率較低的廠商及生產國，進而對經濟部標準檢驗局提出風險評估之建議以節省相關的檢驗人力成本，且降低進口貨品檢驗的時間，同時可以打破現有的抽批方式，提出新的抽批觀念，使本局可以有效的將檢驗能力放在風險高的廠商產品。此外，本研究亦實際探勘經濟部標準檢驗局臺中分局受理的96年及97年之進口廠商(約80間)報驗貨品資料，實驗結果最大可觀察至最大物件組3。在單一物件組的探勘結果中，可發現大部份進口貨品的檢驗皆呈現合格，而以二個物件組來看，發現裕毛屋企業股份有限公司所進口的貨品合格率很高，以進口貨品之生產地來說仍以日本為最多，其次為中國大陸、越南。若以最大物件組(亦即3)來觀察，可以發現夏暉物流有限公司從美國所進口的產品其合格率最高。

關鍵詞：資料探勘、進口商品檢驗、關聯法則、交易資料庫。

1. 前言

我國經濟部標準檢驗局(簡稱標檢局)[1]為確保國人使用市面流通商品之安全與衛生，會針對國內廠商所進口之商品進行抽驗。例如，日前國內所發生之三聚氰胺的事件，便是由本局負責檢驗廠商由國外所進口的商食品或者原料，若檢驗結果不合格時，則限制廠商進口此一貨品。因此，本局在邊境管制貨品上佔非常重要的角色，若管制的好則國人的安全就有保障，若管制的不好對國人的生命財產將會受到相當大的威脅。現行本局的抽批檢驗作業是採取混合抽批，亦即不考慮廠商以往所進口貨品之合格率，以統一機率的方式進行抽檢的動作；或者考量廠商進口某種產品不合格率過高而提高抽批機率，此種作法容易造成現有抽批及檢驗人力無法有

效的發揮，可能造成合格率較高的進口廠商不斷的被抽驗，而合格率低的進口商卻屢次避過抽驗。此外，現行的混合抽批機制亦造成進口貨品的控管品質低落。對照現行關稅局之海關作業原則，我們可以發現他們對於進口績優廠商享受較低的抽驗比率，而將人力放在控管不良廠商上實則似乎是較佳的抽批方式，亦是我們在執行抽批的模式上可以參考的方式。因此，我們提出新的抽批方式來改善現行的抽批模式，希望藉由資料探勘的技術來控管進口產品，而使本局的人力能做最有效的應用。

資料探勘[2, 3]是從大量的資料中，找出隱藏的知識，提供給專業人員做為決策參考。資料探勘[4, 5]主要是運用統計模式、數學演算法、機器學習方法等工具，其目的不只包括收集資料和管理資料，也包含分析和預測的功能。資料探勘和傳統統計學最大的差異，在於前者並沒有預先設計好的假說(Hypothesis)或者假設(Assumption)，也不需要計算其效力(Power)以及p值，因此資料探勘的優點在於可得到新的發現或知識。資料探勘是知識探索找尋隱藏在資料中有用的知識和樣式程序中的一個步驟，資料探勘的資料來源不限於資料庫，也包括其它類型的資訊儲存形式，及範圍廣大的資料倉儲(Data Warehouses)[6]。目前資料探勘常用分析模式有分類、預測、群集、關聯、資料摘要、相依模式、連結分析及順序分析等八類[5]，其中以關聯式規則[7, 8]的探勘方式最常被人運用在實際應用。現行經濟部標檢局的進口貨品抽批機制往往都耗費許多的人力、資源和時間，使得各個廠商們抱怨為何總是檢驗需要這麼久的時間，怎麼總是抽到我的貨物。因此，我們的研究是在不增加現有本局的人力成本條件下，從進口商過去所累積抽驗的合格率的資料上，找出有用的關聯規則，利用這些關聯規則，我們將可以提供更快更精準的抽批檢驗資料。

本論文於第二節介紹相關研究；第三節說明我們的研究步驟以及方法；第四節描述Microsoft SQL 2005中內建的關聯式演算法(Apriori演算法)；第五節則是分析所找出的關聯式規則。第六節則是結論及未來研究

方向。

2. 相關研究

本節將描述資料探勘之關聯法則原理及經濟部標準檢驗局之進口商品報驗程序及相關資料。

2.1. 資料探勘關聯法則原理

關聯規則[7, 8]可以說是計算最簡單的資料探勘演算法，它所應用的原理就是「條件機率」而已。典型的關聯規則形式如下：『假如購買A，則有X%的機率也會同時購買B』。關聯法則有二個重要的指標：一是信心水準(Confidence)；二是支持度(Support)。信心水準[8]顧名思義就是到底這條規則的準確度有多少，從條件機率的公式來看，等於在A的條件下發生B的可能性，因此公式可以利用下式表達：

$$Confidence(A \rightarrow B) = P(B | A) = \frac{P(A \cap B)}{P(A)}$$

信心水準越高，自然這條規則就越有參考價值。但信心水準高固然表示規則具有高準確度，但是否值得轉化為行銷組合呢？所以還必須參考支持度。支持度指的就是符合這條規則的交易次數。支持度[7]的公式如下：

$$Support(A \rightarrow B) = P(A \cap B)$$

信心水準以及支持度不但可以用來判斷規則的有效性之外，同時在演算法的訓練過程中也扮演著非常重要的角色，可以用來大幅縮減規則數量。不過並非所有關聯規則符合顯著性門檻後都有意義，必須再根據關聯規則的隱含資訊價值將規則分成三大類：有用的規則、常識及無法解釋的結果。

Apriori演算法[9]為最早提出關聯式探勘的演算法，其方法特性為簡單、易於實作，它利用不斷地掃描資料庫來重複檢驗候選項目集是否為頻繁項目集，然而當頻繁項目集長度為k時，其掃描資料庫的次數即為k次，

當掃描次數過高時，有可能降低其探勘效率。Apriori演算法包含了二個重要的步驟：(1)反覆的產生候選項目組合搜尋整個資料庫，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。在步驟(1)中，候選項目組的支持度必須大於或是等於使用者所設定的最小支持度門檻值，才能成為大項目組。同樣的，在步驟(2)中所產生的關聯法則其可靠度也必須達到使用者最初設定的最小可靠度門檻值。

在步驟(1)產生大項目組的過程中，Apriori演算法由單一項目組(1-itemset)開始，逐層產生相關項目組。此過程分為二個階段：第一個階段是讀取資料庫一次，找出所有支持度大於最小支持度的項目稱之為 L_k ，其中 k 代表是長度為 k 的大項目集合。第二階段可分為三個步驟：1.利用長度為 $k-1$ 的大項目集合(L_{k-1})來產生候選項目集合(C_k)。2.然後讀取資料庫來計算所有候選項目集合的支持度，將所有支持度大於最小支持度的候選項目集合選出來成為長度 k 的大項目集合(L_k)。3. k 加1，然後重複第2個步驟，直到無法再產生新的候選項目集合為止。

Apriori核心演算法：

```
1   $L_1 = \{\text{large 1-itemsets}\};$ 
2  for( $k=2; L_{k-1}; k++$ ) do begin
3       $C_k = \text{apriori-gen}(L_{k-1});$  //新的
候選集
4      for all transactions in
Database do begin
5           $C_t = \text{subset}(C_k);$ 
6          for all candidates  $C_t$ 
do
7              count++;
8      end;
9       $L_k = \{C_k | \text{count}(\text{minsup})\}$ 
```

```

10  end;
11  Answer= $\bigcup_k L_k$ 

```

以上所列出的即為Apriori演算法，其中的Apriori-gen函式主要的功能就是產生候選項目集，它利用 L_{k-1} 來產生長度為k可能是大項目集合的所有項目。

2.2. 經濟部標檢局之商品報驗程序及資料

目前我國貨物通關之作業流程[1]如圖1所示因所採行之驗放方式不同而有所分別：1.免驗：輸入非銷售之自用品、商業樣品，展覽品或研發測試用物品，其報單單一項次之金額在美金一千元以下或同規格型式之數量未逾商品免驗辦法規定之一者，准予免驗。2.逐批檢驗、監視查驗：實施監視查驗之商品經逐批查驗一定批數符合規定者，採行逐批查核、抽批查驗、書面核放或監視之方式放行。另外針對衛生署委託本局代為實施的食品方面，主要針對產品的特性設定逐批查驗、加強抽批及一般抽批等抽批機率的不同模式來控管進口食品的衛生安全。而不管是商品或食品我們都可以看出抽批方式乃設定一抽批百分比大家在這機率之下共同抽批。當沒有抽中就採書面核放，當有抽中就針對項目進行檢驗，檢驗合格才可以放行。

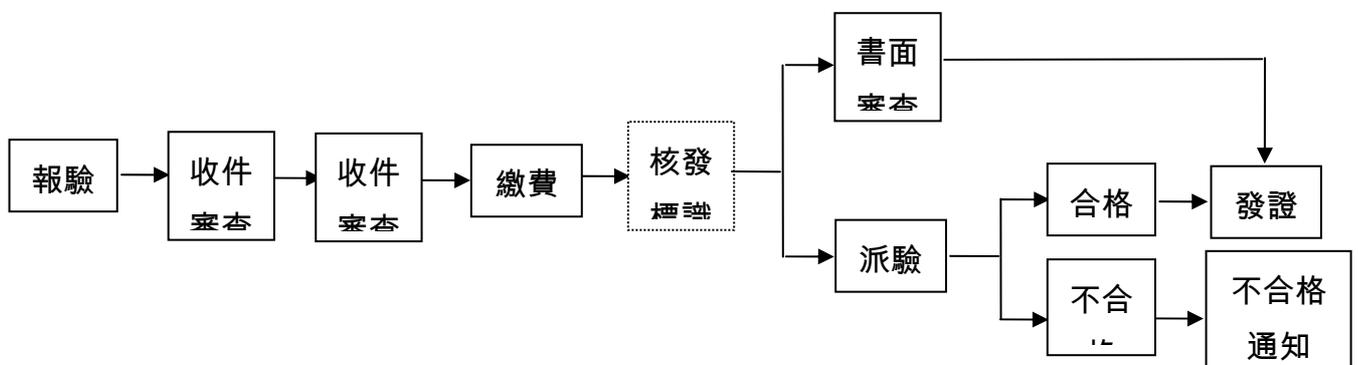


圖1. 標檢局之廠商報驗程序

現今經濟部標檢局的資料庫即為因應我國貨物通關之作業流程，並儲存廠商報驗資料，資料庫內資料種類繁多，包含檢驗申請單的號碼，受理時間，進口日期，報驗義務人(廠商)，廠商住址，連絡電話，進口貨品名稱，製造廠名稱，數量，貨品分類號列，費用，合格與否，經辦人，承辦人及檢驗員。資料庫內所有資料皆存放於同一表格，大致可分為三類，第一類是與標檢局檢驗業務相關；第二類則是與廠商相關；第三類則是與產品相關。其分別描述如下：

1. 標檢局(申請單號碼, 經辦人, 承辦人, 檢驗員, 受理時間, 費用)。
2. 廠商(客戶編號, 廠商住址, 聯絡電話)。
3. 產品(貨品名稱, 貨品編號, 數量, 進口日期, 合格與否, 貨品分類號列)。

在第1項與標檢局相關的資料中，我們可以發現廠商報驗貨品時，必須填寫報驗申請單，並交由標檢局進行登記與處理，此時，標檢局的報驗資料庫中必須產生一個新的申請書號碼，並說明經辦人員、承辦人員及檢驗人員為何人，並註明受理報驗的時間，及欲收取的檢驗處理費用。在第2項與廠商相關的資料中，我們可以發現，曾經報驗過的廠商，標檢局會給予一個證書編號，並說明廠商的住址及聯絡電話等資訊。在第3項與產品相關的資料中，分別有貨品名稱、編碼、報驗物品的數量、進口日期及合格與否等資訊。

3. 研究步驟及方法

本研究是以資料挖掘技術中的關聯規則演算法對標檢局中的進口商報驗資料庫進行挖掘，找出進口廠商、進口貨品、進口國家及其合格與否之間的關連關係。我們的研究步驟及方法如圖2所示。

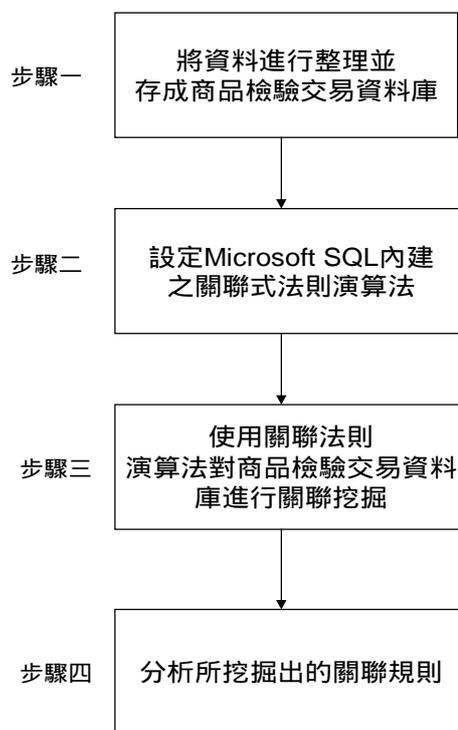


圖2 研究步驟及方法

在步驟一中，我們首先先搜集經濟部標準檢驗局臺中分局中進口商報驗資料庫中的相關資料，並對這些資料進行整理，使其能形成交易資料庫。在進行關聯規則挖掘之前，我們必須在現有的進口商報驗資料庫中，大量搜集進口商在過去進口貨品的項目及合格與否等資訊，或是事先做好一些計算以加速資料挖掘的動作，而且這些轉換及計算的動作只要做一次即可，不必每次進行資料挖掘時都重新再做。因此，在此階段，我們以資料庫設計的原理及概念，對資料進行正規劃設計，使其能儲存於Microsoft SQL 2005資料庫[10]，再利用資料庫中的資料，形成交易資料庫。

在步驟二中，我們設定好Microsoft SQL 2005內建的資料挖掘演算法Apriori。在步驟三中，我們使用步驟二所設定好的演算法對步驟一中的交易資料庫進行資料挖掘。在步驟四中，我們分析我們所挖掘出的結果，並分析比較Apriori在挖掘交易資料庫的效能。

4. 實證分析

本研究以Microsoft SQL Server 2005[10]作為分析處理軟體工具，其提

供許多資料庫應用程式的工具，其中主要包括關聯式資料庫、資料轉換服務、資料庫複寫工作、分析服務套件等技術。在本研究中，我們使用分析服務套件的關聯法則。

4.1. 資料前處理

本研究的資料來源是來自於經濟部標準檢驗局臺中分局中報驗資料庫的資料，其中包含「96年廠商報驗記錄資料表」及「97年廠商報驗記錄資料表」約80家進口廠商進口的相關資料。

在使用SQL Server 2005的關聯法則探勘之前，需先將蒐集來的資料寫入資料庫中，為符合SQL Server程式的演算，因此需設計成可被使用的資料表格。在一般交易的資料表設計當中，都會有主檔與明細檔一對多的表格來存放每次的交易資訊，但本局的檢驗資料皆放於同一資料表中，因此，在本研究中我們只將資料存放在單一資料表中。

4.2. 以關聯法則探勘標檢局之商食品檢驗交易資料庫

Apriori演算法是目前市面上最普遍的關聯規則演算法，針對商食品檢驗交易資料的前置統計量包括：最大規則物件數(Maximum Itemset)、最小支持度(Minimum Support)及最小信賴度(Minimum Confidence)。最大規則物件數是指規則中物件組所包含的最大物件數量。最小支持度是指規則物件或是物件組合必須符合的最低案例數，實驗中我們將支持度設定為0.3。此一數值我們乃參考本局中檢驗能力所設定之值，其可因應本局之需要做適度的調整。最小信賴度則指計算規則所必須符合的最低信心水準門檻，在實驗中設定為2。廠商最少必須報驗的次數至少二次以上才有意義。將商品報驗與一般消費行為作比對，可以將廠商的每次報驗當成一筆交易記錄。Apriori演算法是根據交易資料明細為基礎，遞迴計算候選的物件組以及規則組合，表1為商品檢驗資料庫中的交易，我們將利用此表說明實際運算流程，為求簡單表達，廠商的部份，我們暫以A、B及C表示。

表1. 商品檢驗資料庫之範例交易資料

廠商	品名	進口國	日期	合格
A	鹽水草菇	越南	970120	是
A	鹽水草菇	中國	970317	否
B	鹽水草菇	越南	970317	否
B	鹽水草菇	越南	970317	否
C	葵花子	中國	970426	是
C	葵花子	中國	980109	是
C	食用樹薯澱粉	越南	970109	否
C	食用樹薯澱粉	越南	980109	是
B	食用樹薯澱粉	越南	970109	否
A	食用樹薯澱粉	越南	980109	是

Step 1：首先演算法掃描過全部交易記錄，首先產生交易資料中所包含的所有單一物件數的所有物件組合(ItemSets)，並計算各單一物件組的案例數(支持度)。基本上，根據排列組合公式，如果共有m家廠商報驗n種產品，此種產品有p個進口國家，報驗結果又可分為合格與否2種，就應該產生 $m*n*p*2$ 種組合。以表1為例則會有 $3(A、B及C廠商)*3(鹽水草菇、葵花子及食用樹薯澱粉)*2=18$ 。表2則為首次掃描之結果。

表2. 商品檢驗資料庫之範例首次掃描結果

廠商	次數	品名	次數
A	3	鹽水草菇	4

B	3	葵瓜子	2
C	4	食用樹 薯澱粉	4
進口國	次數	合格與 否	次數
越南	7	是	5
中國	3	否	5

Step 2：如果產品種類很多時，這樣候選物件組數量會非常龐大，同時再向下排列組合時，數量更會是指數的增加，因此參數最小支持度就變得非常重要了。此實驗最小支持度是0.3，那麼支持度不到 $(0.3*10)=3$ 的物件組會予以刪除。這麼做的主要原因在於，規則是利用物件組所構成，如果物件組本身都無法達到最小支持度的門檻時，包含這個物件組的規則就一定也不會達到門檻。因此，在原來產生的物件組中，會把未達支持度門檻的物件組刪除，只保留達到門檻的組合，我們稱之為候選大型物件組(Candidate Large Itemsets)。在這一步，符合候選大型物件組合的組合數只有8組(刪除葵瓜子的組合)。如表3所示。

表3. 商品檢驗資料庫之關聯法則候選物件組1

廠商	次數	品名	次數
A	3	鹽水草 菇	4
B	3	食用樹 薯澱粉	4
C	4		
進口國	次數	合格與 否	次數
越南	7	是	5

中國	3	否	5
----	---	---	---

Step 3：如此一來，單一物件組從10個變成8個。下一步，繼續產生兩個物件的物件組(2-item Itemsets)。根據排列組合原理，如果有n個物件時，應該會產生 C_2^n 總和，那麼套用到此實驗中，就變成 C_2^8 組。因為，物件組是根據底層物件所組成，因此如果物件本身都無法達到最小支持度門檻時，包含這個物件的物件組也一定不可能達到門檻。於是就會根據上一步驟所過濾的物件產生兩個物件的物件組合候選組合，並統計支持度，最後再根據最小支持度進行過濾。將次數大於等於3的篩選出來之後，剩下6組的物件組內容如表4所示。

表4. 關聯法則候選物件組2

物件組合	次數
B、越南	3
C、是	3
鹽水草菇、越南	3
食用樹薯澱粉、越南	4
越南、是	3
越南、否	4

Step 4：一直反覆以上的步驟，直到產出包含「最大規則物件數」+1個物件的物件組為止。此實驗設定規則物件數最多為2，則需要一直迴圈產生物件數為3的物件組為止或是已經沒有可以組合的物件為止。將次數大於等於3且物件組為3的篩選出來，餘下3組的物件組內容如表5。

表5. 關聯法則候選物件組3

物件組合	次數
------	----

B、越南、否	3
--------	---

Step 5：物件組產生會根據「最大規則物件數」+1的門檻中止，或是如果在此步驟中，產生的符合最小支持度物件組數為零時，系統也會中止物件組搜尋。Step 6：將物件組排列組合，計算所有可能的規則，同時計算規則的支持度、信心水準。Step 7：過濾掉不符合最小支持度以及信心水準的規則，此時即完成Apriori演算法的計算過程。

4.3. 實驗結果

4.3.1. 基本工具及參數說明

資料探勘是近年來隨著人工智慧和資料庫技術發展的一門新興技術。他可以從大量資料中，萃取出隱含、過去不為人所知且可信與有效的資訊。以下所列就是使用SQL Server 2005關聯法則[10]提供之參數：

1. MAXIMUM_ITEMSET_COUNT：指定要產生的最大項目集數目。預設值為200000。本實驗亦設定為預設值。
2. MINIMUM_PROBABILITY：指定規則為True的最小機率。例如，將此值設定為0.5是指不產生機率小於50%的規則。本實驗設定為0.3，其可因應本局之需要做適度的調整。
3. OPTIMIZED_IMPORTANCE：規則最小重要性(增益值)門檻，當規則重要性低於此值時則會被刪除。本實驗亦有此一設定。

4.3.2. 實際商品檢驗資料庫結果分析

首先，我們將標檢局台中分局的部分報驗資料檔(Excel檔案)先匯入SQL 2005的資料庫中，如圖3所示。



圖3. 將標檢局資料匯入SQL 2005

4.3.2.1. 以關聯法則探勘的流程

建置關聯規則的步驟如下：

Step 1：在資料來源檢視中，加入實際交易的資料檔，同時輸入友善命名。
如圖4所示。

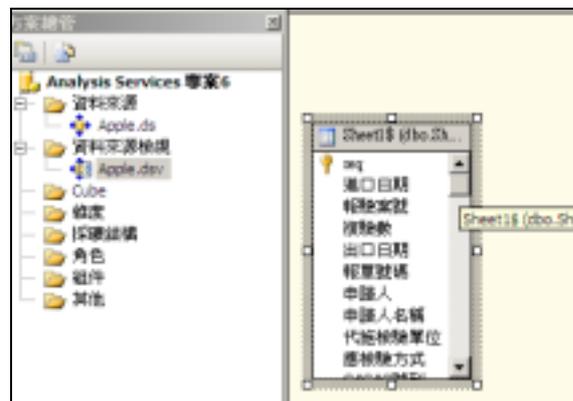


圖4. SQL 2005建置資料來源檢視

Step 2：將滑鼠移到「方案總管」的「資料結構」圖示按右鍵，選取「新增採礦結構」。

Step 3：精靈會詢問要使用關連式資料庫或只是多維度分析資料來源。在這裡我們選擇關聯式資料庫。如圖5所示。

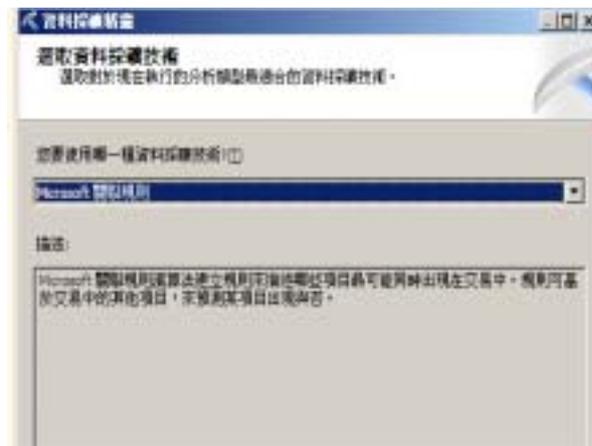


圖5. SQL 2005探勘選擇畫面

Step 4：在接下來的畫面中選取「Microsoft關聯法則演算法」。

Step 5：選取方才設好的資料來源顯示。

Step 6：設定「案例資料表」與「巢狀資料表」。由於本次實驗只用單一表格來表示標檢局的進口貨品報驗資料，因此只要選擇案例資料表。如圖6所示。



圖6. SQL 2005選擇資料表類型

Step 7：變數設定。報驗號碼設定成索引值，而廠商、產品名稱、進口國家及檢驗結果設定成「輸入」及「可預測」。如圖7所示。

Step 10：處理資料探勘模型即可完成關聯法則。

4.3.2.2. 96年廠商報驗記錄探勘結果與分析

圖10資料取標檢局之廠商報驗記錄做關聯法則探勘，不同參數所探勘出來的資料。我們以支持度為0.1進行探勘所得，因此，最小支持度之物件組數最少應該263，而最大物件組數為3。

序號	大小	描述
201	1	檢驗結果=T
202	1	中國人志願-裕毛屋企業股份有限公司
207	2	中國人志願-裕毛屋企業股份有限公司, 檢驗結果=T
219	1	生產地-日本
228	2	生產地-日本, 中國人志願-裕毛屋企業股份有限公司
238	2	生產地-日本, 檢驗結果=T
257	3	生產地-日本, 中國人志願-裕毛屋企業股份有限公司, 檢驗結果=T
402	1	中國人志願-夏暉物流有限公司
403	2	中國人志願-夏暉物流有限公司, 檢驗結果=T
349	1	生產地-越南
357	1	生產地-美國
358	2	生產地-越南, 檢驗結果=T
322	2	生產地-美國, 檢驗結果=T
296	1	生產地-美國
294	1	生產地-中國大陸
309	2	生產地-中國大陸, 檢驗結果=T
300	2	生產地-美國, 檢驗結果=T
309	2	生產地-美國, 中國人志願-夏暉物流有限公司
205	3	生產地-美國, 中國人志願-夏暉物流有限公司, 檢驗結果=T

圖10. 96年廠商報驗記錄探勘結果

由圖10，我們可以得知96年廠商報驗的資料中，單一物件組中，以進口廠商及生產地具有最大的關聯性，顯示以中部地區來看進口貨品的進口廠商，以裕毛屋企業股份有限公司為最大量，而生產地以日本為最多，其次為中國大陸、越南。若觀察二個物件組，則可以發現印證單一物件組所探勘之結果，另外，亦可發現來自日本之進口貨品合格率最高。若以最大物件組(亦即3)來觀察，可以發現夏暉物流有限公司從美國所進口的產品其合格率最高。因此，由上面探勘結果，我們可以將檢驗人力放在其他進口商所進口的產品，在進行檢驗抽批機制時，能有更好的作法。

4.3.2.3. 97年廠商報驗記錄探勘結果與分析

圖11資料取標檢局之廠商報驗記錄做關聯法則探勘，不同參數所探勘出來的資料。我們以支持度為0.1進行探勘所得，因此，最小支持度之物件組數最少應該282，而最大物件組數為3。

序號	大小	項目數
367	1	檢驗結果=T
718	1	中國人合資-裕毛屋企業股份有限公司
717	2	中國人合資-裕毛屋企業股份有限公司, 檢驗結果=T
719	1	生產地=日本
718	2	生產地=日本, 中國人合資-裕毛屋企業股份有限公司
718	2	生產地=日本, 檢驗結果=T
717	2	生產地=日本, 中國人合資-裕毛屋企業股份有限公司, 檢驗結果=T
402	1	中國人合資-夏暉物流有限公司
401	2	中國人合資-夏暉物流有限公司, 檢驗結果=T
349	1	生產地=美國
338	1	生產地=美國
330	2	生產地=美國, 檢驗結果=T
334	2	生產地=美國, 檢驗結果=T
296	1	生產地=美國
294	1	生產地=中國大陸
295	2	生產地=中國大陸, 檢驗結果=T
288	2	生產地=美國, 檢驗結果=T

圖11. 97年廠商報驗記錄探勘結果

由圖11，我們可以得知97年廠商報驗的資料中，單一物件組中，我們可以發現大部份進口貨品的檢驗皆呈現合格，而以二個物件組來看，發現裕毛屋企業股份有限公司所進口的貨品合格率很高，以進口貨品之生產地來說仍以日本為最多，其次為中國大陸、越南。若以最大物件組(亦即3)來觀察，可以發現夏暉物流有限公司從美國所進口的產品其合格率最高。因此，由上面探勘結果，我們可以將檢驗人力放在其他進口商所進口的產品，在進行檢驗抽批機制時，能有更好的作法。

4.3.2.4 96年及97年廠商報驗記錄探勘結果與分析

圖12資料取標檢局之廠商報驗記錄做關聯法則探勘，不同參數所探勘出來的資料。我們以支持度為0.1進行探勘所得，因此，最小支持度之物件組數最少應該549，而最大物件組數為3。

由圖12，我們可以得知96及97年廠商報驗的資料中，單一物件組中，我們可以發現大部份進口貨品的檢驗皆呈現合格，而以二個物件組來看，發現裕毛屋企業股份有限公司所進口的貨品合格率很高，以進口貨品之生產地來說仍以日本為最多，其次為中國大陸及越南。若以最大物件組(亦即3)來觀察，可以發現夏暉物流有限公司從美國所進口的產品其合格率最高。因此，由上面探勘結果，我們可以參考關稅局的作法將這些優良廠商

降低抽批機率，而將檢驗人力、物力放在高風險的廠商、產地及產品上，使本局的人力能有效的利用。

序號	次數	大小	項目集
5281	1		檢驗結果 = Y
1436	1		申請人名稱 = 裕毛屋企業股份有限公司
1454	2		申請人名稱 = 裕毛屋企業股份有限公司, 檢驗結果 = Y
1438	1		生產地 = 日本
1436	2		生產地 = 日本, 申請人名稱 = 裕毛屋企業股份有限公司
1436	2		生產地 = 日本, 檢驗結果 = Y
1434	3		生產地 = 日本, 申請人名稱 = 裕毛屋企業股份有限公司, 檢驗結果 = Y
964	1		申請人名稱 = 夏暉物流有限公司
800	2		申請人名稱 = 夏暉物流有限公司, 檢驗結果 = Y
698	1		生產地 = 越南
675	1		生產地 = 美國
660	2		生產地 = 越南, 檢驗結果 = Y
647	2		生產地 = 美國, 檢驗結果 = Y
590	1		生產地 = 泰國
588	1		生產地 = 中國大陸
578	2		生產地 = 中國大陸, 檢驗結果 = Y
568	2		生產地 = 泰國, 檢驗結果 = Y

圖12. 96及97年廠商報驗記錄探勘結果

5. 結論與未來研究方向

本研究使用資料探勘之關聯法則技術探勘經濟部標檢局臺中分局之96年及97年廠商進口量較大的80間廠商報驗資料庫。關聯法則也稱購物籃分析，通常用於銷售記錄中商品與商品之間的關聯。在經濟部標檢局的廠商報驗資料庫中，我們同樣可以用此方法找出廠商、進口產品、進口國家及合格與否之間的關聯規則。我們從單一物件組至物件組3分別探討其關聯關係(包含分析廠商、產品、進口國家及合格與否)，並針對探勘結果預測未來廠商來進行報驗時，應該採取何種的抽驗方式。透過我們的實驗我們可以了解到那些進口商是時常進口不合格產品，那些進口商所進口的產品合格率較高。對於時常進口不合格的廠商，本局應該考慮將檢驗能力放在這些廠商所進口的產品上，以達到控管產品的目的。在我們的實驗中，我們發現96年及97年中部地區裕毛屋是國內目前最大的進口廠商，此外，我們亦可判斷，夏暉物流有限公司是家有品質的進口商，其進口貨品的合格率很高。對於優良廠商所進口的產品，我們可以參考海關作法降低抽驗比率。我們的未來研究方向是可以加入其它屬性探討彼此間的關聯關係，或者增加全局的資料庫，增加可用性。由於本局的實際報驗資料庫中仍有

許多屬性(例如原料、進口數量等可以加入探討其關聯法則)，因此，可以考慮下一步將更多的資料帶入關聯法則。

6. 參考文獻

- [1]. 經濟部標準檢驗局 . Available from: <http://www.bsmi.gov.tw/wSite/index.jsp>.
- [2]. S. Chaudhuri and U. Dayal, *An Overview of Data Warehousing and OLAP Technology*. ACM SIGMOD Record, 1997. **26**: p. pp. 65-74.
- [3]. J. Hipp, U. Güntzer, and G. Gakhaeizadeh, *Algorithms for Association Rule Mining - A General Survey and Comparison*. SIGKDD Explorations, 2000. **2**(2): p. pp.1-58.
- [4]. J.S. Pork, P.S. Chen, and P.S. Yu, *An Effective Hash based Algorithm for Mining Association Rules*. ACM SIGMOD, 1995: p. pp. 175-186.
- [5]. P.N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. 2006: Addison-Wesley.
- [6]. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. 2000, San Francisco, CA: USA:Morgan Kaufmann.
- [7]. R. Agrawal, T. Imielinski, and A. Swami. *Mining Association Rules Between Sets of Items in Large Databases*. in *In Proceeding of SIGMOD Conference*. 1993.
- [8]. R. Agrawal and R. Srikant. *Fast Algorithms for Mining Association Rules*. in *In Proceedings of the 20th VLDB Conference*. 1994.
- [9]. J. Han, et al., *Mining Frequent Patterns without Candidate Generation*. Data Mining and Knowledge Discovery 2004. **8**: p. pp.53-87.
- [10]. 尹相志, *SQL Server 2005 Data Mining 資料採礦與 Office 2007 資料採礦增集*. 2007.