

# 利用關聯法則分析個股之間的關聯性

黃有評

國立臺北科技大學

電機工程系

yphuang@ntut.edu.tw

李家政、謝尚琳

大同大學

資訊工程研究所

e9406002@ms.ttu.edu.tw

## 摘要

長久以來許多研究專注在金融市場的關聯性，嘗試在金融市場去找出或是預測可靠依據的原則，提供相關的訊息或是預測結果來供投資者或是決策者參考，並且在金融市場中挖掘一些不容易發現且含有意義的資訊。本研究利用資料探勘中關聯法則的技術，來探討台灣股市交易市場中有相互影響的公司是否有連動關聯性，以提供投資人作為判斷投資的一個依據。本研究主要是利用關聯法則在台灣的股市交易市場中去探勘出標的公司中有關聯性的公司有哪些？針對電子類股中的3百多家上櫃上市公司中依據人力資源網站所做的公司排名分類標的公司來分析在區段交易日中的關聯性為何，達到提供投資人一個不錯的依據。研究中針對標的個股中的每日交易做跌幅的切割，在支持度為20%且信賴度50%的實驗結果中，發現在2007整年度的資料範圍內可探勘約400條規則為最多，在2007Q2Q3的資料範圍內約40條符合具有意義的規則，在這些有意義的規則下搭配視覺化的分析下，所顯示出的公司都具有相當的關聯性。

**關鍵詞：**資料探勘、關聯法則、股市關聯。

## 1. 序論

### 1.1. 研究背景與動機

以目前台灣股市交易市場中，上市上櫃公司已有數百家之多，股市預測一直都是被列入研究的主題[12-16]，大部分是以技術指標以及線型資料搭配不同的預測模型達到預測效果，但是台灣股市交易市場的技術指標數值常因政治因素以及外資法人的炒作下而受影響，因此如何利用系統化之計算方式，試圖找出上市上櫃中個股公司之間的關聯程度為何，使得投資人可藉由分析出來的結果加以判斷是否適合投資與否，以及建立良好的投資風險管理。

因此，本研究將不在於股價預測，而是利用基本的每日交易資料去找出哪些個股與個股之間是有較高的關聯性，在這些關聯性中找

出非一般常識能歸納的個股關聯法則，由於台灣股市交易上市上櫃的公司數量過多，因此資料探勘(Data Mining)就是一種可以運用在股市個股分析的方法。

因此，利用資料探勘中常被使用的關聯法則(Association Rules)技術，找出個股之間的連動性，讓資訊效率化、正確化，使得投資者可以依據探勘出的規則作出正確的投資決策。

### 1.2. 研究目的

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自動化發現一些有用資訊的流程，並萃取出隱含、過去不為人知且可信與有效的知識。也可以說是依據使用者所設定的參數，在一群未經處理的資料中找到使用者感興趣的資訊，經過某些特殊處理後，做為決策者判斷的參考依據。而關聯法則是其中常被使用的一種方法。關聯法則常被用來分析表示交易記錄中顧客所購買產品間的關聯性，以做為產品陳設、促銷或決策用[4,7-9]。

關聯法則在尋找產品間的關聯性有強大的能力，我們將需要分析個股的每日交易漲跌設定為商品集(Item Set)，商品數量會經由不同的幅度切割有不同的數量，將每月、季、半年、一年、二年的交易日設定為交易集(Transactions Set)。而股市交易數量一年約有280天的交易，由於公司的獲利產品會因為週期關係而與不同的公司有關聯性，故不適合拿超過3年的連續交易資料分析，因此本研究將只針對三年內的股市交易資料作為探勘的基礎，找出個股之間的關聯性。因此所探勘出的關聯性更可以提供專業投資決策者的判斷依據，更可減少投資者分析的時間與成本。

本篇論文共有五個章節，第一節緒論說明研究動機與目的；第二節為文獻探討，介紹資料探勘與關聯法則；第三節為本篇論文的研究方法，說明用關聯法則所分析的股市方式；第四節則是實驗結果分析；第五節則為結論和研究建議，說明本研究的實驗結論並為後續研究提出建議。

## 2. 文獻探討

### 2.1. 資料探勘

隨著網際網路的發達以及資料庫技術的成熟，人們取得資料變得非常的容易，再加上許多網際網路的應用其實就是一個自動化的資料收集工具，資料量之大已幾乎接近爆炸的程度。雖然我們所收集的資料愈來愈多，然而在這一堆令人眼花瞭亂的資料當中，究竟隱藏了哪些真正有用的資訊或規則？如何將這些有用的資訊或規則從大量的資料當中挖掘出來，作為輔助決策的參考，便成為我們所關心的焦點[2]。

資料探勘是一種能夠在大量資料中自動化發現一些有用資訊的流程，它可以用來處理大量資料，而且可以發現一些新奇以及有用的樣式，而這些知識是之前所未知的，另外，資料探勘技術也在於預測未來。但並非所有資訊發掘的工作都稱為資料探勘，例如在資料庫管理系統中找出個人紀錄、或是經由網路搜尋引擎來找出特定的網頁等，這些工作都只是所謂的資料檢索；雖然這類的工作都很重要，而且也許可加入一些複雜的演算法以及資料結構，但是他們都是依賴於傳統的電腦科學技術，同是利用資料的明顯特徵來建立索引結構以增進組織的效率與快速的搜尋。儘管如此，資料探勘技術事實上也常被用來強化資訊檢索系統的效率和搜尋的結果。

資料探勘只是從資料庫裡發掘知識過程中的一個部份而已，因為整個部份包含了如何將原始資料轉換成有用的資訊。這個過程包含了一系列從資料的前處理一直到資料探勘結果的後處理的轉換過程[3]。

### 2.2. 關聯法則

關聯法則可以說是資料探勘過程中很重要的演算法，其實它所應用的原理就是大家熟悉的條件機率概念。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』

我們利用一個簡單的圖形(圖 1)來說明這條規則是如何計算出來的。假設在某段期間中，購買 A 商品的顧客共計有 80 名，而購買 B 商品的人共計有 100 人，而兩者的交集同時購買 A 以及 B 的顧客，也就是圖形中的重疊區域共計有 65 人。

此時，在購買 A 商品的 80 個客戶中有 65 人也買了 B 商品；而在購買 B 商品的 100 人中，有 65 人也同時購買了 A 商品，因此，我們可

以計算出來兩條關聯規則：

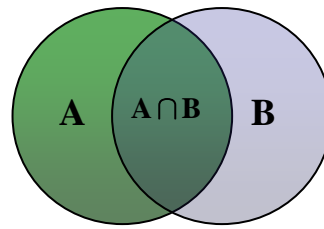


圖 1、條件機率示意圖。

- 『假如購買 A，則有 81.25%的機率也會同時購買 B(65/80=0.8125)』
- 『假如購買 B，則有 65.00%的機率也會同時購買 A(65/100=0.65)』

然而在這麼多的規則中，我們要如何判斷哪些規則對我們來說是有用的，我們可以從兩個指標來著手：

#### A. 信賴度(confidence)：

信賴度顧名思義就是到底這條規則的準確度有多少，從條件機率的公式看來，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示：

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

以上述例子來說，購買 A 也會去買 B 的信心水準為 81.25%，信心水準越高，自然這條規則就越有參考價值。

#### B. 支持度(support)：

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？所以還要參考支持度。支持度指的就是符合這條規則的交易次數。支持度的公式如下：

$$\text{Support}(A \rightarrow B) = P(A \cap B)$$

信心水準以及支援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[1]。

### 2.3. 關聯分析

滿足最小支持度與最小信心水準不一定保證可以提供具體有用的資訊。觀察下面的例子

- 假設在包含 10000 筆交易紀錄的資料庫中，有 6000 筆交易包含碳粉匣，7500 筆交易包含報表紙，並且有 4000 筆交易同時包含碳粉匣和報表紙。假設最小支持度為 30% 最小信心水準為 60%。下列的關聯

法則會產生誤導的作用：碳粉匣→報表紙  
【支持度=40%,信賴度=67%】

這是因為購買報表紙的機率為 75%，大於信賴度 67%，但是購買碳粉匣反而降低購買報表紙的可能性。基本上，資料探勘的結果只是提供我們可能的重要資訊，企業必須根據資訊制定決策，再透過行動產生價值。若實際效益和預期結果反差較大時，則必須修正相關參數，重新進行下一個循環的探勘工作[2]。

我們可以利用計算項目集之間的相關程度來預測法則的使用效益。項目集 A 和項目集 B 的相關程度計算方式如下：

$$Lift = \frac{P(A \cap B)}{P(A)} - P(B) > d$$

- 若  $-1 \leq Lift < 0$ ，表示 A 和 B 是負關聯，亦即 A 的出現會造成 B 出現的機率降低。
- 若  $0 < Lift \leq 1$ ，表示 A 和 B 是正關聯，亦即 A 的出現會造成 B 出現的機率增高。
- 若  $Lift = 0$ ，表示 A 和 B 是無關聯，亦即 A 和 B 是無關的。

當相關程度大於 0 時，表示使用法則的效果會比較顯著。但是當相關程度小於 0 時，則表示使用法則的效果可能不佳。

### 3. 研究方法

前面章節已經介紹探勘方法以及關聯法則與分析，本章節將利用之前介紹的方式來探勘股市交易市場中個股之間的關聯性。

#### 3.1. 方法架構

本論文的方法架構如圖 2 所示，其中程序為：

步驟一：資料蒐集。包含從 2006 年第三季至 2008 年第二季台灣證券交易所提供的每日交易資料、Yahoo 股市交易資訊中的每家上櫃上市公司的營收項目比例以及網路人力銀行所提供的百大公司排名列表。

步驟二：篩選需要探勘公司的交易資料。將所收集到的每日交易資料搭配百大公司排名列表去篩選出哪些公司需要探勘。

步驟三：將篩選過後的每日交易資料做前置處理。由於台灣證券交易所提供的每日交易資料中並沒有我們所需要的唯一因子「漲跌幅」，因此需透過轉換將交易資料轉換成漲跌幅以及將漲跌幅做切割成五種類別。

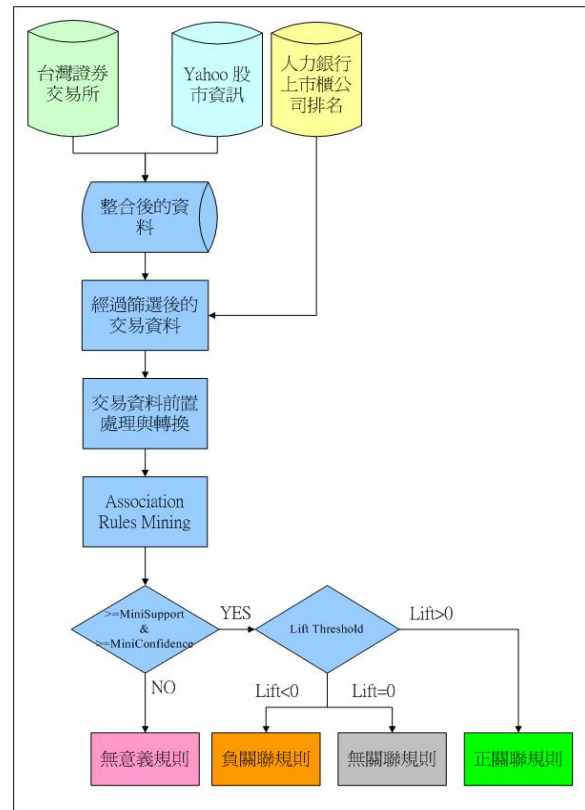


圖 2、實驗架構。

步驟四：利用 Apriori 演算法探勘規則以及分析關聯規則。

#### 3.2. 資料數據的前置處理

台灣證券交易所提供的每日交易資料中含有證券代號、證券名稱、成交股數、成交筆數、成交金額、開盤價、最高價、最低價、收盤價、+/-漲跌價、最後揭示買價、買進揭示委託數量、最後揭示賣價、賣出揭示委託數量等資料，但是我們探勘中需要的唯一因子「漲跌幅」，所以我們將利用原有的資訊去做前置處理的轉換將「+/-漲跌價」、「收盤價」轉換成我們所需要的「漲跌幅」，轉換的計算方式如下[12]：

$$Range = \frac{\pm P(c)}{P(e) \mp P(c)} * 100\%$$

Range：定義為漲跌幅

P(c)：定義為漲跌價

P(e)：定義為收盤價

#### 3.3. 資料的切割

由於台灣股市的漲跌幅是有最高與最低的限制分別為 7%、-7%，在本研究中將定義五種漲跌幅的類別分別為下列所示[5,10-11]。

$Range > 3.5 \in bUp$

$0.01 < Range \leq 3.5 \in sUp$

$-0.01 \leq Range \leq 0.01 \in None$

$-3.5 \leq Range < -0.01 \in sDn$

$Range < -3.5 \in bDn$

bUp：定義為大漲

sUp：定義為小漲

None：定義為平盤

sDn：定義為小跌

bDn：定義為大跌

### 3.4. 資料來源的選擇

在本研究中我們在台灣證券交易所中的電子工業類股裏有上市上櫃中的三百多家電子公司搭配網路人力銀行所提供的公司排行中屬於筆記型電腦以及記憶體生產的公司選擇了二十九家公司，並且選擇六個階段的每日交易資料當做研究的資料輸入來源，分別如表 1 及表 2 所示[6]：

表 1、篩選探勘之相關公司代號。

以篩選過之相關公司列表		
TW.2317	TW.2324	TW.2331
TW.2354	TW.2356	TW.2362
TW.2377	TW.2382	TW.2384
TW.2451	TW.2474	TW.3005
TW.3701	TW.6145	TW.6235
TW.2344	TW.2357	TW.2408
TW.3017	TW.3481	TW.2348
TW.2364	TW.2421	TW.3231
TW.3533	TW.2353	TW.2376
TW.2425	TW.3474	

表 2、交易區間。

各階段每日交易區間	
來源	交易日期區間
Source1	2007 整年度
Source2	2006Q3Q4 ~ 2008Q1Q2
Source3	2007Q3Q4 ~ 2008Q1Q2
Source4	2007Q1Q2
Source5	2007Q2Q3
Source6	2007Q3Q4

由於電子公司所生產的產品可能會因為技術上的影響，可能造成每年所生產的主力將有所改變，因此在本研究中我們在分析公司之間的關聯性上將採取最長探勘的時間區間為三年，最少為每季。

## 4. 實驗結果

本章節將針對六種不同交易日數量區間的資料經過關聯法則探勘後的結果加以分析。

### 4.1. Apriori 演算法參數說明

表 3 是在本研究中 Apriori 演算法所使用的參數說明。

表 3、Apriori 參數設定。

參數	說明	實驗所使用門檻值
minimal support	指定演算法產生規則之前必須包含項目集的最小案例數目。將此值設定為小於 1，是以總案例數的百分比來指定最小案例數目。將此值設定為大於 1 的整數，是以必須包含項目集的絕對案例數目來指定最小案例數目。	本實驗所設定值為 20%
minimal confidence	指定項目集可支援的最大案例數目。如果此值小於 1，則此值代表總案例數的百分比。大於 1 的值代表可包含項目集的絕對案例數目。	本實驗所設定值為 50%
Lift value print	指定演算法在產生規則的同時也將相關的 Lift 增益值列印出來。	本實驗所設定值為 Enable
minimal item set	指定項目集內所允許的最小項目數目。	本實驗所設定值為 1
maximal item set	指定項目集內所允許的最大項目數目。	本實驗所設定值為 5

### 4.2. 不同交易日數量探勘結果與分析

在該實驗中，分別先後探勘六種不同階段

的交易資料去探勘出相關的規則，在以 Confidence $\geq$ 60% 和 Lift $\geq$ 20% 參數設定為門檻分類出有意義的規則搭配個股日線圖加以比較分析後再藉由 Yahoo 網站所提供的各家公司營收項目表分析出更有關聯性的公司。

#### 4.2.1. 2007 年整年探勘結果與分析

2007 年整年度交易日數量為 247 天，探勘結果符合 miniSupport 和 miniConfidence 的規則數量為 402 條規則，而符合 Confidence 和 Lift 的規則數量為 16 條規則，將分類為有正相關規則，而符合 Confidence 但是 Lift 為 0 的規則數量為 4 條並分類為無相關規則，Lift 小於 0 的規則數量為 41 條並分類為負相關規則。

在這個資料來源中可以發現該實驗結果所產生的規則較多，主要的原因是因為台灣股市的上市上櫃的公司行號所上市日期不一定，所以會導致在資料中會有很多不相關的規則，但是可以透過 Lift 的門檻去消除這些無意義的規則，由於要保留每家公司每日交易資料的完整性，故在資料選擇時不去刪除，圖 3 為其中有相關聯規則之公司日線比較。

#### 4.2.2. 2006Q3Q4 ~ 2008Q1Q2 探勘結果與分析

2006 年第三季至 2008 年第二季整體交易日數量為 494 天，探勘結果符合 miniSupport 和 miniConfidence 的規則數量為 160 條規則，而符合 Confidence 和 Lift 的規則數量為 5 條，將分類為有正相關規則，在這兩年交易資料中根據所設定的門檻中並未發現不相關規則和負相關規則，圖 4 為其中有相關聯規則之公司日線比較。

#### 4.2.3. 2007Q3Q4 ~ 2008Q1Q2 探勘結果與分析

2007 年整年度交易日數量為 267 天，探勘結果符合 miniSupport 和 miniConfidence 的規則數量為 122 條，而符合 Confidence 和 Lift 的規則數量為 20 條，將分類為有正相關規則，在 2007 年度中根據所設定的門檻中並未發現不相關規則和負相關規則，圖 5 為其中有相關聯規則之公司日線比較。

#### 4.2.4. 2007Q1Q2 探勘結果與分析

2007 年前半年度交易日數量為 119 天，探勘結果符合 miniSupport 和 miniConfidence 的規則數量為 298 條，而符合 Confidence 和 Lift 的規則數量為 25 條，將分類為有正相關規則，

在 2007 年度中根據所設定的門檻中並未發現不相關規則和負相關規則，圖 6 為其中有相關聯規則之公司日線比較。

#### 4.2.5. 2007Q2Q3 探勘結果與分析

2007 年 Q2Q3 半年度交易日數量為 125 天，探勘結果符合 miniSupport 和 miniConfidence 的規則數量為 295 條，而符合 Confidence 和 Lift 的規則數量為 45 條，將分類為有正相關規則，而符合 Confidence 但是 Lift 為 0 的規則數量為 2 條並分類為無相關規則，Lift 小於 0 的規則數量為 24 條並分類為負相關規則，圖 7 為其中有相關聯規則之公司日線比較。

#### 4.2.6. 2007Q3Q4 探勘結果與分析

2007 年下半年度交易日數量為 128 天，探勘結果符合 miniSupport 和 miniConfidence 的規則數量為 323 條，而符合 Confidence 和 Lift 的規則數量為 23 條，將分類為有正相關規則，而符合 Confidence 但是 Lift 為 0 的規則數量為 3 條並分類為不相關規則，Lift 小於 0 的規則數量為 32 條並分類為負相關規則，圖 8 為其中有相關聯規則之公司日線比較。

#### 4.3. 不同交易日數量總結分析

我們在實驗中分別實驗了六種不同交易數量的資料來源，以及二十九間屬於筆記型電腦或是記憶體上游到下游的公司，在下列表格可以清楚看到他們所探勘出來的規則數量，以及符合條件門檻的設定值所能夠參考的規則數量，在實驗結果中可以分別看出當實驗的交易日數量超過 1 年以上所探勘出來的規則都會比較少量且符合門檻的數量也會比較少，這也符合了在電子業中的公司的關聯性可能會依據每年的技術能力或是趨勢而有所變化。

在表 4 中 S1 的資料來源範圍中，雖然規則數量多，但是大部分的規則都是因上市櫃的日期所導致產生大量不相關規則，但是在 Lift 的門檻值設定下可以將這些無意義的規則先行排除，圖 9~11 為表 4 的直方圖。

本研究除了在支持度大於 20% 與信賴度大於 50% 的實驗外，也針對固定支持度變動的信賴度以及固定信賴度搭配變動支持度下實驗。

在圖 12、圖 13、圖 14、圖 15 中分別可以看到探勘出的結果數量，且可以發現當 Support 值越低，探勘的結果數量越大，相對的不相關規則與負相關規則的數量也隨之增

加，雖然可以探勘出更多負相關規則但是也造成了正相關規則大量的產生導致不容易分析。

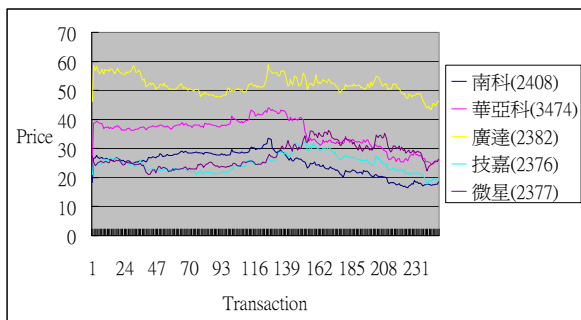


圖 3、S1 關聯公司日線比較。

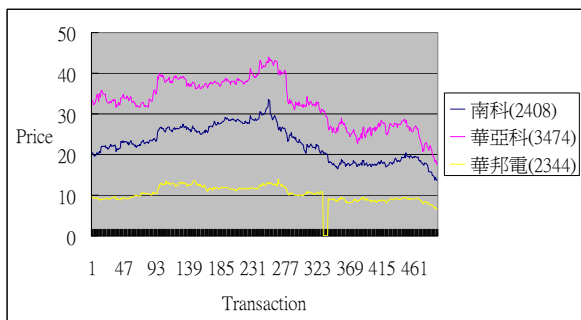


圖 4、S2 關聯公司日線比較。

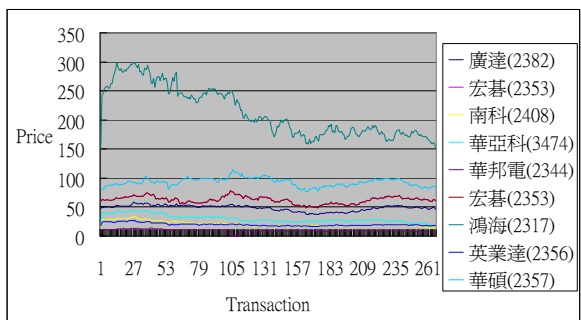


圖 5、S3 關聯公司日線比較。

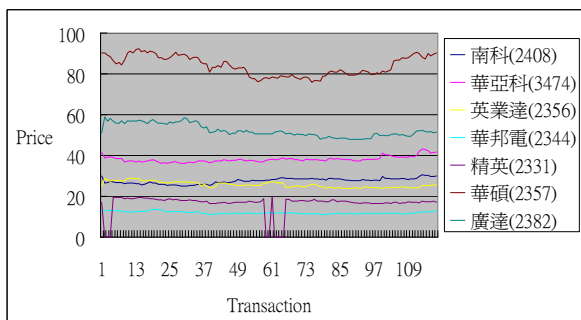


圖 6、S4 關聯公司日線比較。

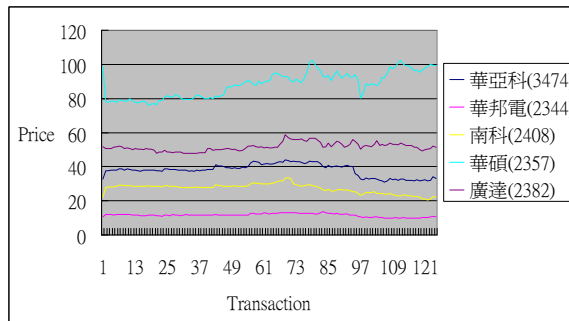


圖 7、S5 關聯公司日線比較。

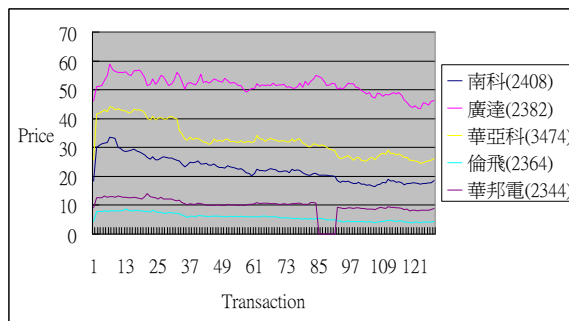


圖 8、S6 關聯公司日線比較。

表 4、實驗結果。

Source	S1	S2
資料範圍	07'	06'Q3Q4~08'QQ1Q2
Supp./Confid.	0.2/0.5	0.2/0.5
規則數量	<b>406</b>	160
符合 Lift	16	5
不符合 Lift	<b>341</b>	155
Lift=0	<b>4</b>	0
Lift<0	<b>41</b>	0
Source	S3	S4
資料範圍	07'Q3Q4~08'QQ1Q2	07'Q1Q2
Supp./Confid.	0.2/0.5	0.2/0.5
規則數量	122	298
符合 Lift	20	25
不符合 Lift	102	273
Lift=0	0	0
Lift<0	0	0
Source	S5	S6
資料範圍	07'Q2Q3	07'Q3Q4
Supp./Confid.	0.2/0.5	0.2/0.5
規則數量	295	322
符合 Lift	<b>44</b>	23
不符合 Lift	225	264
Lift=0	2	3
Lift<0	24	32

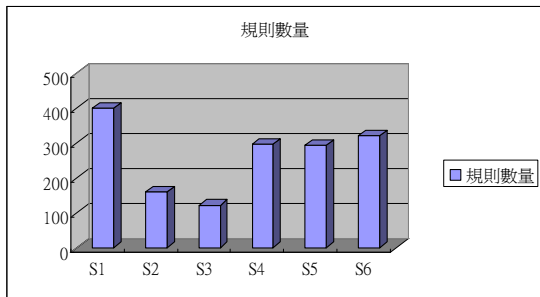


圖 9、規則數量。

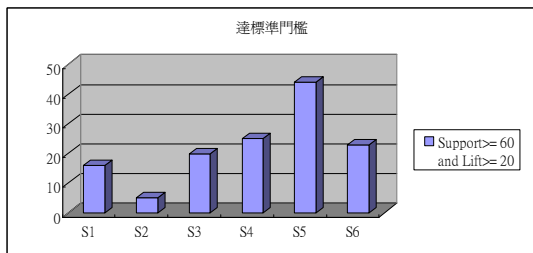


圖 10、符合標準規則。

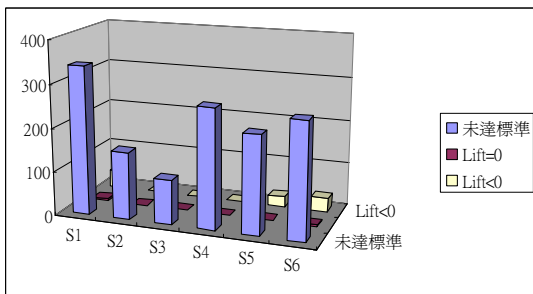


圖 11、Lift 相關結果數量比較。

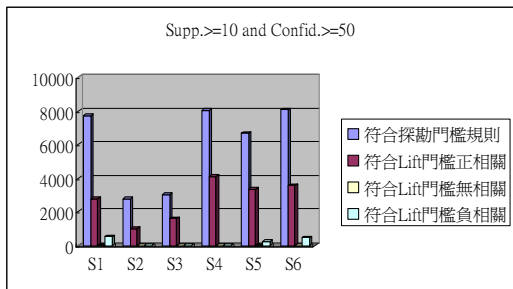


圖 12、Supp.  $\geq 10\%$  and Confid.  $\geq 50\%$ 。

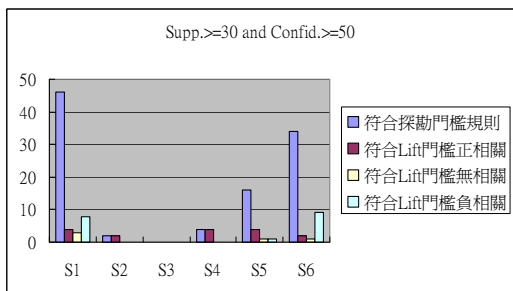


圖 13、Supp.  $\geq 30\%$  and Confid.  $\geq 50\%$ 。

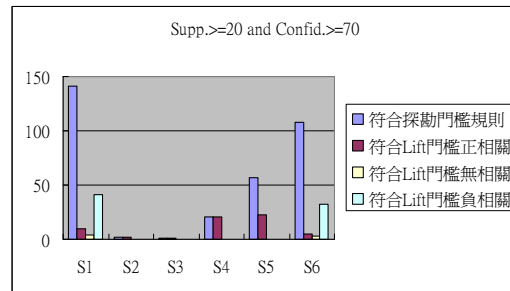


圖 14、Supp.  $\geq 20\%$  and Confid.  $\geq 70\%$ 。

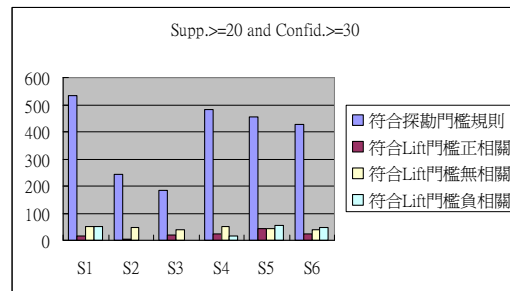


圖 15、Supp.  $\geq 20\%$  and Confid.  $\geq 30\%$ 。

## 5. 結論與建議

在金融市場中如何探勘出有價值且有意義的規則來代表公司之間的關聯性是不容易的一件工作，在本研究中的實驗數據中可以發現資料探勘中的關聯法則不僅僅可以用來分析購物籃的購買狀況分析，亦可以應用在許多關聯性分析上，實驗數據中雖然無法根據探勘出來的數據直接說明探勘出的規則中公司之間的關聯性，但是搭配公司的交易日線圖的視覺化顯示比較可以發現許多規則是可以看出他們的關聯性。

在本研究中在探勘分析中加入 Lift 的增益當作一個門檻限制的條件，可以提高規則是否較符合興趣且有意義，在實驗結果中，可以看出 Lift 的門檻可以較容易分辨出那些無相關或是負相關程度的規則。以下為 Lift 的優缺點。  
**優點：**較容易明確指示出有關聯、無關聯、負關聯的成度表示。

**缺點：**在門檻的部份比較難以去明確需要設定多少，因為門檻值的設定可能會錯失可能更有興趣的規則。

在股市資料中雖然有龐大的歷史資料，但是如何在這些資料中選擇需要探勘的交易量大小是值得去改善的，以及網路人力銀行的公司間的分類和排行如何去做一個適當選擇，這些因素的改善都有機會發掘更多有意義的規則來提供給使用者與決策者一個重要的參考。

## 誌謝

本論文承國科會計畫編號 NSC97-2221-E-027-034-MY3 之補助得以順利完成，特此誌謝。

## 參考文獻

- [1] 尹相志, *SQL Server 2005 Data Mining 資料探礦與 Office 2007 資料探礦增益集*, 第 176-183 頁, 2007。
- [2] 曾憲雄、蔡秀滿、蘇東興、曾秋蓉, *資料探礦(Data Mining)*, 2008。
- [3] 施雅月、賴錦慧, *資料探礦(Introduction to Data Mining)*, 2008。
- [4] 宋文傑, *評估台灣通訊產業的財務績效—多變量區別分析法及灰關聯度整體性分析法之應用*, 東華大學國際經濟研究所碩士論文, 2003。
- [5] 陳吉相, *使用分段式前置處理的關聯規則挖掘來分析國際股票指數*, 逢甲大學資訊電機工程研究所碩士論文, 2005。
- [6] 謝育龍, *快速的跨交易資料挖掘之研究*, 逢甲大學資訊電機工程研究所碩士論文, 2006。
- [7] 邱秋婷, *屬性導向方法應用於證卷交易相對關係規則之挖掘*, 中央大學資訊管理研究所碩士論文, 2007。
- [8] 饒美倫, *使用關聯法則預測與分析台灣樂透彩*, 大同大學資訊工程研究所論文, 2007。
- [9] X.D Wu, C.G Zhang and S.C. Zhang, "Efficient mining of both positive and negative association rules," *ACM Trans. on Information Systems*, vol. 22, no. 3, pp.381-405, July 2004.
- [10] R. Rastogi and K. Shim, "Mining optimized association rules with categorical and numeric attributes," *IEEE Trans. on Knowledge and Data Engineering*, vol. 14, no. 1, pp.29-50, Jan.-Feb. 2002.
- [11] S. Brin, R. Rastogi and K. Shim, "Mining optimized gain rules for numeric attributes," *IEEE Trans. on Knowledge and Data Engineering*, vol. 15, no. 2, pp.324-338, March-April 2003.
- [12] Y.K. Kwon, S.S. Choi and B.R. Moon, "Stock prediction based on financial correlation," in *Proc. of Genetic and Evolutionary Computation Conf.*, Washington DC, USA, pp.2061-2066, June 2005.
- [13] T. Cagatay and B. Badur, "The similarity of stock prices and indices in progressing information technology environment," in *Proc. of PICMET*, Portland, Oregon, USA, pp.944-950, August 2007.
- [14] V. Sehgal and C. Song, "SOPS: Stock prediction using web sentiment," in *Proc. of Seventh IEEE Int. Conf. on Data Mining*, Omaha, NE, pp.21-26, Oct. 2007.
- [15] Z. Zheng and H.W. Wang, "An efficient association rule mining algorithm and business application," in *Proc. of Int. Conf. on Communications, Circuits and Systems*, Kokura, pp.959-965, July 2007.
- [16] M.A. Mittermayer, "Forecasting intraday stock price trends with text mining techniques," in *Proc. of the 37<sup>th</sup> Annual Hawaii Int. Conf. on System Sciences*, Hawaii, pp.1-10, Jan. 2004.