

上市公司間股價漲跌之關聯與預測－關聯 探勘之研究

ASSOCIATION AND FORECAST OF THE RISE AND DROP OF STOCK PRICE BETWEEN THE COMPANIES – ASSOCIATION MINING APPROACH

葉怡成 林文盟

中華大學資訊管理學系

I-Cheng Yeh Wen-Meng Lin

Department of Information Management

Chung-Hua University

摘要

本研究從股價漲跌記錄中以關聯分析去找出關聯規則。結果顯示，關聯規則存在於這些資料之中，並且這些規則可以用來導出產業關聯。此外，從漲跌記錄產生的關聯規則可預測股價漲跌。當漲、跌被定義為 1% 的價格變化時，統計結果顯示，這些規則可以預測下跌趨勢。也就是說，當這些記錄產生關聯規則：「股票 A 下跌 → 股票 B 下跌」，此規則顯示當股票 A 股價下跌時，股票 B 有很高的機率在當日下跌。若股票 B 當天股價持平時，股票 B 隔日有很高的機率會下跌。

關鍵字：資料探勘、關聯規則、股市分析、產業關聯

ABSTRACT

This research studies association rules from stock price movements. The results discovered that the association rules do exist in the data and can help to

draw the industrial relationships. In addition, the association rules can be used to forecast future stock price changes. When rise and drop are defined as 1% change of price, the statistical results showed, these rules may forecast price drop tendency. In other words, when these data produced the association rule: "When Stock A drops, then Stock B will drop on the same day." This rule appears that when the price of stock A drops, Stock B has a very high probability to drop on the same day, and when the price of Stock B didn't drop on the same day, Stock B will have a very high probability to drop next day.

Keyword: data mining, association rule, stock price, industry relationship, forecast

壹、簡介

二十世紀中期以後，科技與經濟的發展突飛猛進，經濟市場的蓬勃發展，隨之造就了股票交易市場的興盛。各種大大小小的企業無不力求上市、上櫃，期使公司的獲利能更上一層樓。而投資者也藉著這個機會，將本身的資金投入其中，心中無不掛念著公司股票의 漲跌。也因為如此，股票投資學也就受到更多人的注目與研究，各門各派的學說也就如雨後春筍般的冒出芽來，且不論是以哪種角度來探討股市分析，均有其支持者的存在。在台灣，上市的公司就高達 700 家左右，而上櫃的公司也高達 500 家左右。投資者如果能在進行投資決策時，考量到這近 1200 家公司彼此的影響力，或是藉由其中的產業結構去找到特殊的影響力，可能可以增加獲利。一般來說，投資者在進行買賣股票時，通常會由三個面向去決定投資決策：(1) 技術面分析：根據股價變動的趨勢，找尋買賣股票最佳的時點。(2) 基本面分析：根據國內外經濟情勢、個別產業前景、公司經營績效等資訊，評估股票真實的價值，做為進出股市及選擇個股的依據。(3) 消息面分析：根據媒體或公司發布的財經新聞，預測投資此公司的獲利性。

以往投資者根據上述的三個面向去判斷投資的標的是否是最佳的目標與最佳的投資時間點為何，但效果常常是非常有限的。這可能是因為股市的行為十分複雜，很難用簡單的方法就去找出一套非常有效的投資準則。

近年來，資料探勘 (Data mining) 技術的發展，許多研究者嘗試著使用資料探勘

的技術去進行投資決策的分析，試圖找出一套有效的投資準則。雖然資料探勘的技術有非常多種類，但以往主要都是利用分類探勘與迴歸探勘去建立股市預測模型，很少利用關聯探勘來進行股市的分析（蕭正南，1998）。關聯探勘定義為給予一組足夠數量的紀錄，其中每筆紀錄都登記了一些項目，而可以從中找出一個能夠以某些項目出現與否來預測其它項目出現與否的相依規則。因此本研究旨在利用關聯探勘來進行台灣股市股價漲跌的關聯分析。

本研究與文獻中的研究方法相較之差異及貢獻有：

- 一、關聯規則與產業關聯圖：以關聯規則產生之產業關聯圖探討產業關聯。
- 二、關聯規則之統計檢定：檢定表 1 的九個假說。
- 三、關聯規則在股價漲跌之預測：探討以關聯規則預測股價漲跌的可行性。

貳、文獻回顧

一、關聯分析

關聯探勘可定義為：「給予一組記錄，每筆記錄登記了一些項目。找出一個能夠以某些項目出現與否來預測其它項目出現與否的關聯規則。」

關聯分析有以下重要參數：

- (一)支持度 (Support)：所謂的支持度是某一特定關聯出現在關聯規則中的頻率。
- (二)信賴度 (Confidence)：當情況一出現時，發生情況二的機率在關聯規則中稱為信賴度。即關聯規則 $A \rightarrow B$ 的 Confidence (B|A) 可表示如下：

$$Confidence(B|A) = \frac{Support(A \cup B)}{Support(A)} \quad (1)$$

關聯分析定義的數學式表達為：「找出資料集中所有支持度 $\geq \text{min_support}$ ，且信賴度 $\geq \text{min_confidence}$ 的關聯規則。」關聯分析的演算法相當複雜，但可以在很多已出版文獻中找到，故在此不加贅述，可參考 Agrawal, Imielinski, and Swami (1993)，Agrawal and Srikant (1994) 等文獻。

表 1 假說檢定

假說	
1.	依據關聯規則預測當天應該「下跌（上漲）」，而「未下跌（上漲）」者，隔天會發生「下跌（上漲）」的機率大於「上漲（下跌）」的機率。
2.	依據關聯規則預測當天應該「上漲」，而「未上漲」者，隔天會發生「上漲」的機率大於「下跌」的機率。
3.	依據關聯規則預測當天應該「下跌」，而「未下跌」者，隔天會發生「下跌」的機率大於上漲的機率。
4.	依據關聯規則預測當天應該「下跌（上漲）」，而「持平」者，隔天會發生「下跌（上漲）」的機率大於「上漲（下跌）」的機率。
5.	依據關聯規則預測當天應該「下跌（上漲）」，而「上漲（下跌）」者，隔天會發生「下跌（上漲）」的機率大於「上漲（下跌）」的機率。
6.	依據關聯規則預測當天應該「上漲」，而「持平」者，隔天會發生「上漲」的機率大於「下跌」的機率。
7.	依據關聯規則預測當天應該「上漲」，而「下跌」者，隔天會發生「上漲」的機率大於「下跌」的機率。
8.	依據關聯規則預測當天應該「下跌」，而「持平」者，隔天會發生「下跌」的機率大於「上漲」的機率。
9.	依據關聯規則預測當天應該「下跌」，而「上漲」者，隔天會發生「下跌」的機率大於「上漲」的機率。

在零售業中，每次交易即是一筆記錄，每種購買的貨品便是一個項目。由於在零售業條碼已被普遍使用，因此收集交易資料非常容易。例如：

1. 買了鐵錘、拔釘器的交易事件中，有 80% 買了鐵釘，便是一條關聯規則：

$\{\text{鐵錘, 拔釘器}\} \rightarrow \{\text{鐵釘}\}$ 信賴度=0.80

2. 修車廠需要零件 A、B 與 C 的修車事件中，有 60% 也需零件 D 與 E，得關聯規則：

$\{\text{A, B, C}\} \rightarrow \{\text{D, E}\}$ 信賴度=0.60

這些項目不一定全是交易項目，也可以是交易的情境（時間、地點）。當買主是持會員卡或信用卡時，項目也可以是顧客資料（年齡、性別、所得）。例如：

1. 一家發行會員卡的量販店從消費者過去的消費行為資料庫中發覺，在週五來買啤酒的年輕男性顧客通常（65%）也會買一包嬰兒尿布。便是一條關聯規則：

{週五，年輕，男性，啤酒}→{嬰兒尿布} 信賴度=0.65

因此，這家量販店在賣啤酒貨架旁放置了嬰兒尿布，這使得啤酒與嬰兒尿布的銷售量都增加了。

關聯分析的演算法可參考文獻 (Han, Cai, & Cercone, 1992; Miller & Yang, 1997)。

二、關聯分析與股市的投資分析

關聯探勘可找出從一些看似不相關，但卻存在某種特殊關聯的規則。因此有學者將此方法應用在找出股價表現關聯上，例如：Tung, Lu, Han, and Feng (1999) 提出一有別於傳統的關聯探勘演算法，用以探勘具有時間序列關係的交易資料，例如股市上市公司收盤價資料隨著時間變化會有不同。利用修改過的 FITI (First Intra Then Inter) 演算法，探勘出股市交易市場中如 A 公司和 B 公司同一天上漲，則 C 公司有 X% 的機率在同一天上漲之關聯規則。

股市的交易收盤價是屬於連續值資料，進行漲跌分析後，可獲得上漲與下跌的百分比資料，若以此進行關聯探勘，必將耗費大量資源與時間去進行規則的挖掘，勢必得將資料進行離散化，將漲跌百分比區分為數個區塊，因此 Miller and Yang (1997) 提出了將分析的資料進行離散化後，再去進行關聯探勘，研究結果證實此方法可以快速且有效的去挖掘出關聯規則。

Lu, Han, and Feng (1998) 亦發現現行傳統的關聯探勘演算法無法挖掘出類似 A 公司和 B 公司同一天上漲後，C 公司會在之後某一天有 X% 的機會上漲。因此修改了 Apriori 演算法，使之可以處理跨天數項目集的關聯規則問題，並將此演算法稱為 EH-Apriori。之後利用 1996 年新加坡股市每日收盤價的資料進行關聯探勘，並定義當漲跌幅度超過 3% 才被分類為上漲或下跌，若不滿足 3% 的門檻值，則不進行任何分類，亦即不考慮持平的問題，接著將上漲與下跌的資料分開進行關聯探勘。結果顯示此方法確實可以從下跌類別與上漲類別中，分別找出有效的關聯規則。因為上述 EH-Apriori 方法不將上漲與下跌的資料放在一起，因此並無法找出如果同一天 A 公司上漲與 B 公司上漲，隔一天 C 公司會下跌這類型的關聯規則。

蕭正南 (1998) 利用由 Han et al. (1992) 等人所提出的 AOI (Attribute-Oriented

Induction) 資料探勘的方法，先對所收集到的股票相關資料進行概念階層化，將資料轉換成一株株的分類樹。接著再透過關聯規則搜尋方法把與股價漲跌有關的規則找出來。再利用所蒐集到的資料和產生出來的規則配合，將未來可能的股價的上下限找出，供投資人參考。並以電子類股中的光寶電子、聯華電子、宏碁電腦、大眾電腦及台積電等五家作為實驗樣本，以此研究所提出的方法進行股價的分析與預測，證實所找出的關聯規則能有效提升投資效益。

參、方法與結果

本研究的進行步驟如下：

1. 建立漲跌資料：以台灣證券交易所公佈的上市公司股價之每日收盤價作為分析資料，並將此資料存成一矩陣表格。以適當的漲跌幅度作為參數，將所收集的資料進行轉換，以建立一包含每日哪些漲或跌之個股資料表，並整理成適當之格式。
2. 建立關聯規則：根據所收集的資料量大小，設定適當的關聯探勘參數，包含支持度、信賴度。透過關聯探勘產生隱含在漲跌資料中的關聯規則。
3. 分析關聯規則：分析所產生的關聯規則中所隱含的上下游、競爭、同盟等產業關聯。

一、建立漲跌資料

技術面分析中所使用的資料大多與股價相關，而觀察一家公司營運的趨勢，最容易使用的資料即為其每日收盤價之漲跌。因此我們在此收集台灣證券交易所內所有上市公司的原始股價資料，收集的時間範圍為 2005 年全年度每日收盤價。而由所收集的資料顯示，2005 年一共有 246 個交易日。

在獲得原始股價資料後，接著決定每日之漲跌門檻，做為當日屬於漲或跌的依據。因為如果只將股價變化分成上漲及下跌二種，可能會造成不易建立正確的關聯規則，例如某股上漲 0.1% 與另股上漲 0.1% 根本不存在關聯性，如二者都被視為「上漲」反而對找到真正的關聯而言是一種雜訊。必須是強烈的上漲或下跌才是有意義的樣本。反之，被視為「上漲」或「下跌」的樣本太少，對找到真正的關聯而言也是一種障礙。因此，原則上希望能達到每家公司平均一年出現 25% 左右屬於上漲及下跌。實驗過後所得數據如表 2，由此可知，漲跌門檻以 1% 較為適宜，此門檻可使每家公司平均一年出現上漲的次數為 56.6 次 (23%)，下跌的次數為 64.1 次 (26%)，持平的次數

表 2 漲跌幅與項目數比較

漲跌幅 (%)	±0	±0.5	±1	±1.5	±2
平均上漲數目	99.7	78.7	56.6	41.9	31.5
平均下跌數目	111.6	89.5	64.1	45.1	32
持平數目	29.9	72.9	120.3	154.0	177.5
上漲出現比例	41%	32%	23%	17%	13%
下跌出現比例	45%	36%	26%	18%	13%

為 120.3 次 (51%)。

當一天之中有超過 50% 的上市公司皆超過我們所設定的漲跌幅時，很可能是由其它外在的重大利空、利多衝擊造成的。為了降低此因素影響關聯規則的正確性，我們將當日上漲與下跌的股票總和超過所有股票總數一半的資料刪除，因此交易日由 246 日減少為 218 日，樣本數為 218 乘以上市股票個數。

二、建立關聯規則

首先，調整支持度門檻與頻集內的最大項目數門檻，並將信賴度門檻設為 60%，藉此來觀察關聯規則數目的變化。結果如圖 1 所示，可知當頻集內最大項目數門檻定為 4 以上時，所得到的關聯規則數目不會改變。因為頻集內的最大項目數與信賴度門檻無關，而與支持度門檻有關，支持度門檻越低，頻集內的最大項目數越高。由於頻集內的最大項目數門檻大於 4 時與等於 4 時有相同數量的關聯規則，可見其門檻值定為 4 就足夠。因此，以下將頻集內最大項目個數設定為 4，接著進行下一個實驗。

接著，調整支持度門檻與信賴度門檻，以期獲得適當數量的關聯規則供使用者進行分析。原則上，支持度門檻越低，產生的頻集越多，因此會有一些支持度不算太高的頻集也因通過門檻而產生，由這些頻集所產生的關聯規則可靠度就不足。其次，信賴度門檻越低，產生的關聯規則越多，因此會有一些信賴度不算太高的關聯規則也因通過門檻而產生，這些關聯規則因信賴度不高，用處不大。在關聯分析中，並不存在一個決定多少條關聯規則是最佳的理論，而視關聯規則的使用目的而定。由於在本研究中關聯規則的使用目的在於預測股價漲跌，支持度太低可能會造成產生的關聯規則可靠度低，故支持度門檻不宜太低，但由圖 2 可知，支持度門檻設定為 19 以上時，關聯規則的數目急速下降；而關聯規則太少會造成無法對各公司股價間的關聯作全面

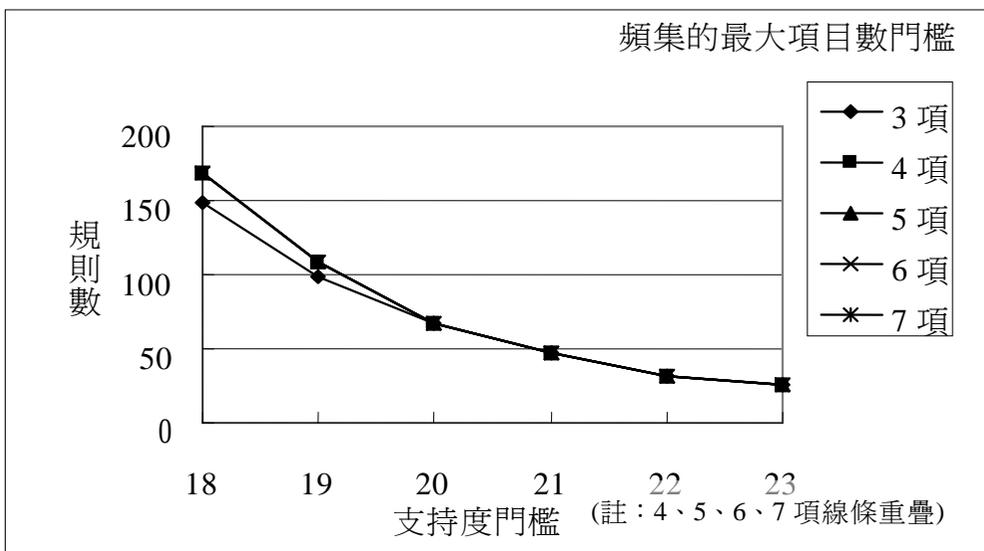


圖 1 最大項目數門檻與支持度門檻變化產生關聯規則數量圖

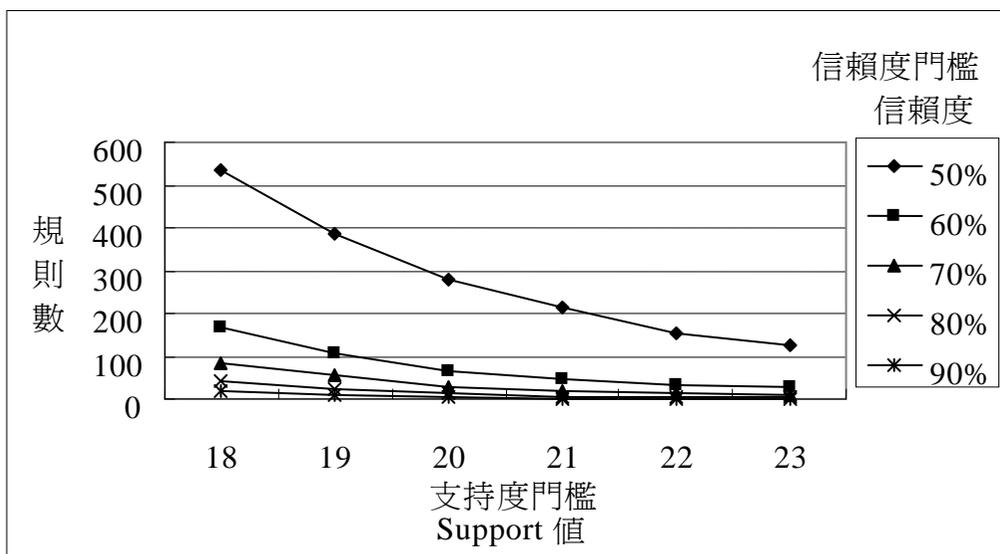


圖 2 信賴度門檻與支持度門檻變化產生關聯規則數量圖

性的分析。此外，信賴度低的關聯規則對預測股價漲跌的價值低。故本研究支持度門檻設定為 18，信賴度門檻設定為 60%，此時獲得 168 條關聯規則。

三、分析關聯規則

表 3 為頻集『國碩跌、利碟跌、精碟跌』所產生的關聯規則。此頻集的支持度為 20，一共產生三組關聯規則。分別為：

- (一)「利碟跌，精碟跌→國碩跌」，信賴度為 0.87，表示當股票市場發生利碟和精碟同時下跌，則有 87%國碩也會一併下跌。
- (二)「國碩跌，精碟跌→利碟跌」，信賴度為 0.769，表示當國碩與精碟同時下跌時，利碟有 76.9%的機率會一起下跌。
- (三)「國碩跌，利碟跌→精碟跌」，信賴度為 0.833，表當國碩和利碟同一天下跌時，精碟也會在同一天下跌的機率有 83.3%。

此頻集中三家上市公司皆為光碟片生產廠商，其所產生的關聯皆為發生同時下跌。在同業的競爭下，可能發生的情況有四種：(1) 甲廠商有獲得大筆訂單的利多消息，股價上漲；其他廠商則因沒搶到訂單，造成市場看空，股價下跌。(2) 整體產業前景看好，投資者看好此一產業，造成此產業的股票同時上漲。(3) 整體產業景氣低迷，投資者不看好此一產業，造成此產業的股價同時下跌。(4) 彼此雖然為競爭同業，但是彼此客戶群穩定，生產能力也不足以吃下整個市場，此時彼此的股價並沒有明顯的互相影響。因為在資料期間，光碟片生產的利潤過低，擁有智財的公司索取過高的權利金，導致光碟片產業股價下滑，屬於上述的第三種情況，因此產生此組頻集甚為合理。

表 4 是將頻集屬於「液晶面板業」的挑出，一共有三組頻集，其中僅缺少「彩晶」這家公司，就剛好是「面板五虎」了。而這三組頻集所產生的五條關聯規則前項與後項皆為上漲的狀態，可以得知在面板產業上，這四家公司在股價上漲方面是相當一致的。投資者在進行投資行為時，就可以參考此一關聯規則來進行決策。

表 5 為光學鏡頭業所產生之關聯規則。此兩組頻集中包含了「大立光、亞光、今國光」等三家公司，此三家公司皆為此產業的領導廠商。且因 2005 年的數位相機與照相手機使用的光學鏡頭需求大增，因此相關的產業皆呈現蓬勃發展的趨勢，股價的表現也相當的理想。因此產生同漲的關聯規則。

表 3 國碩跌、利碟跌和精碟跌之關聯規則

No.	前項	後項	信賴度
104	2443 利碟 DOWN; 2396 精碟 DOWN	→ 2406 國碩 DOWN	0.87
105	2406 國碩 DOWN; 2396 精碟 DOWN	→ 2443 利碟 DOWN	0.769
106	2406 國碩 DOWN; 2443 利碟 DOWN	→ 2396 精碟 DOWN	0.833

表 4 面板產業之關聯規則

頻集：3009 奇美電 UP；3012 廣輝 UP（支持度為 20）

No.	前項	後項	信賴度
17	3012 廣輝 UP	→ 3009 奇美電 UP	0.625
18	3009 奇美電 UP	→ 3012 廣輝 UP	0.606

頻集：3009 奇美電 UP；2409 友達 UP（支持度為 19）

No.	前項	後項	信賴度
14	2409 友達 UP	→ 3009 奇美電 UP	0.679

頻集：3012 廣輝 UP；2475 華映 UP（支持度為 20）

No.	前項	後項	信賴度
17	2475 華映 UP	→ 3012 廣輝 UP	0.69
18	3012 廣輝 UP	→ 2475 華映 UP	0.625

表 5 光學鏡頭業之關聯規則

頻集：3008 大立光 UP；3019 亞光 UP（支持度為 22）

No.	前項		後項	信賴度
17	3019 亞光 UP	→	3008 大立光 UP	0.647

頻集：3019 亞光 UP；6209 今國光 UP（支持度為 21）

No.	前項		後項	信賴度
14	3019 亞光 UP	→	6209 今國光 UP	0.618

四、關聯規則與產業結構之關係－關聯圖之應用

爲了讓投資者能較清楚的瞭解關聯規則中產業的關聯性，此節以 SQL Server 2005 繪製本章所產生的 168 條關聯規則之關聯圖。結果如圖 3 所示。可發現相同產業的公司股價常常發生同漲或同跌的情況，並被關聯圖繪在一起，形成化工、光學鏡頭、光碟、IC 設計、IC 測試、石英元件、液晶面板、飯店娛樂、鋼鐵、營建等產業群聚。例如圖中可發現到鋼鐵業同漲的公司與同跌的公司各形成一小部份關聯圖，可判斷鋼鐵業在股價漲跌表現很容易與同產業的公司相似，因此分別形成上漲與下跌的關聯圖。而在鋼鐵業附近可以看到營建業的關聯圖，此兩種產業在營運上又有相當的影響，營建業往往需要大量的鋼材來建築房子或各種公共建設，當某一產業的股價發生波動時，另一個產業就很容易受到波及，造成彼此股價的影響。

肆、以關聯規則預測股價漲跌之可行性分析

關聯規則具有方向性，當前項指向後項之關聯規則成立時，後項指向前項之關聯規則並不一定會成立。亦即，當我們得到一組關聯規則爲 $B \rightarrow A$ 時，僅能說明當 B 漲或跌時，A 經常也會同時漲或跌；並不能反推回 $A \rightarrow B$ ，也就是無法說明當 A 發生漲或跌時，B 也會隨之同時漲或跌。圖 4 的示意圖可以說明此一觀念：

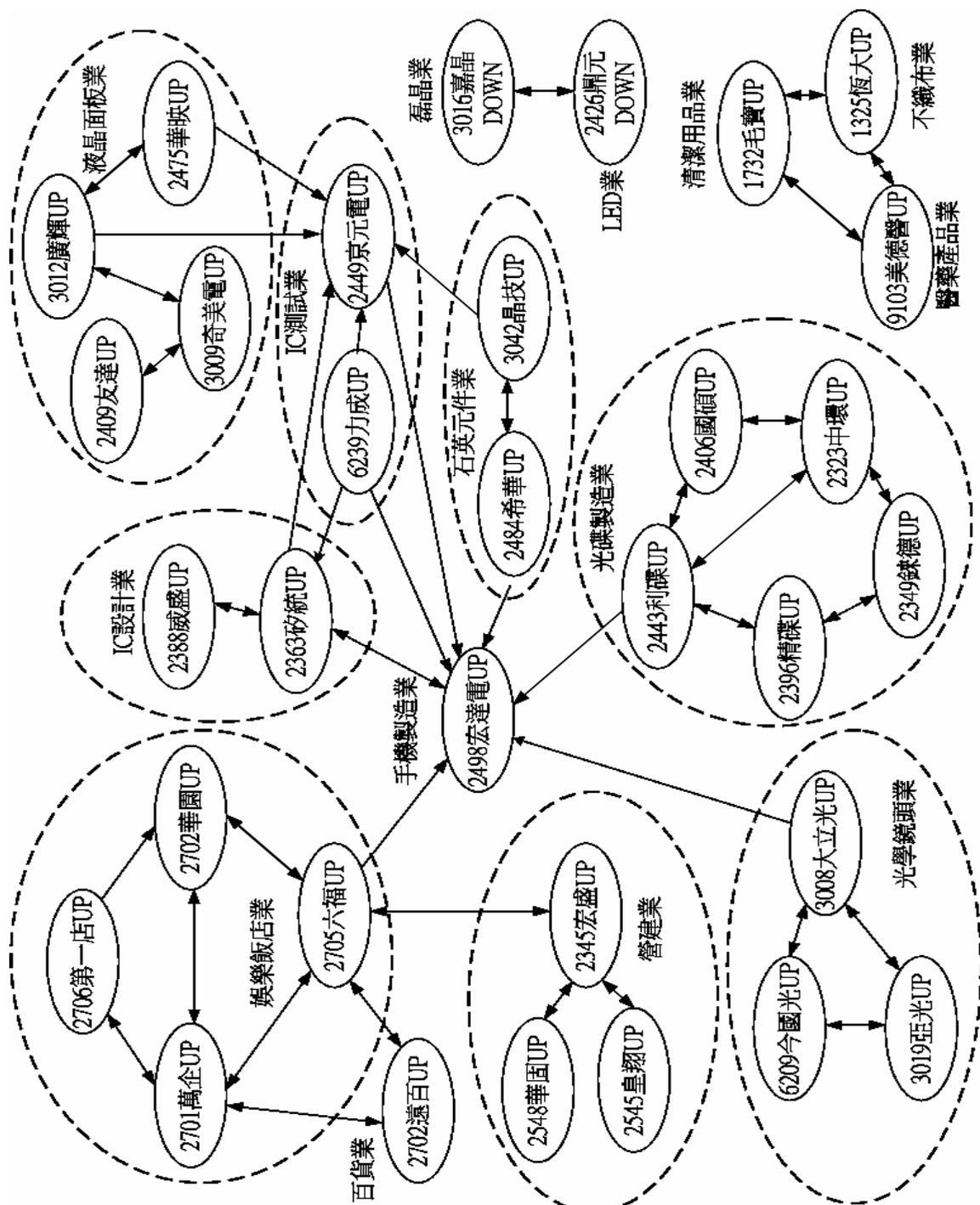


圖 3 媒體關係圖

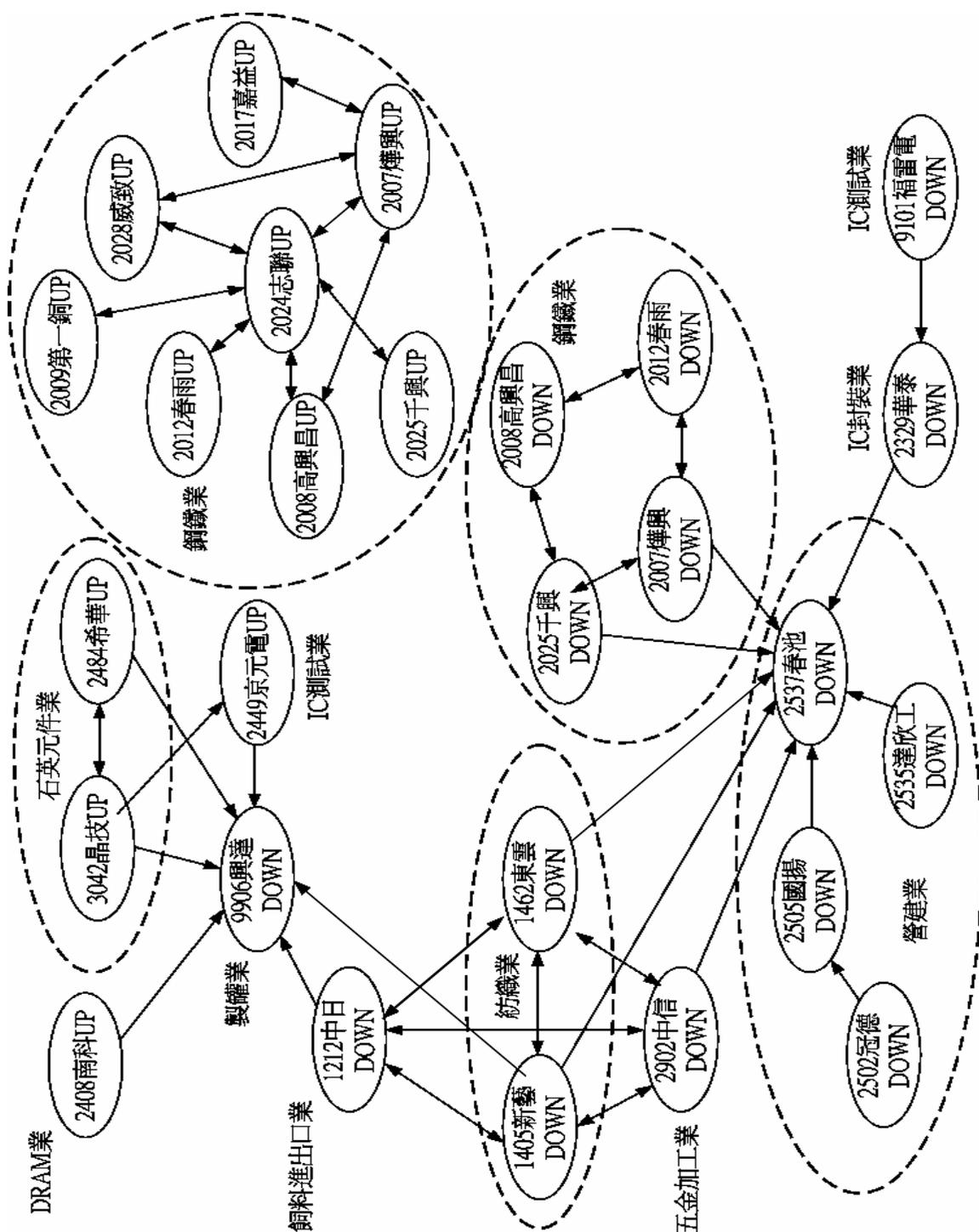


圖 3 整體關聯圖 (續)

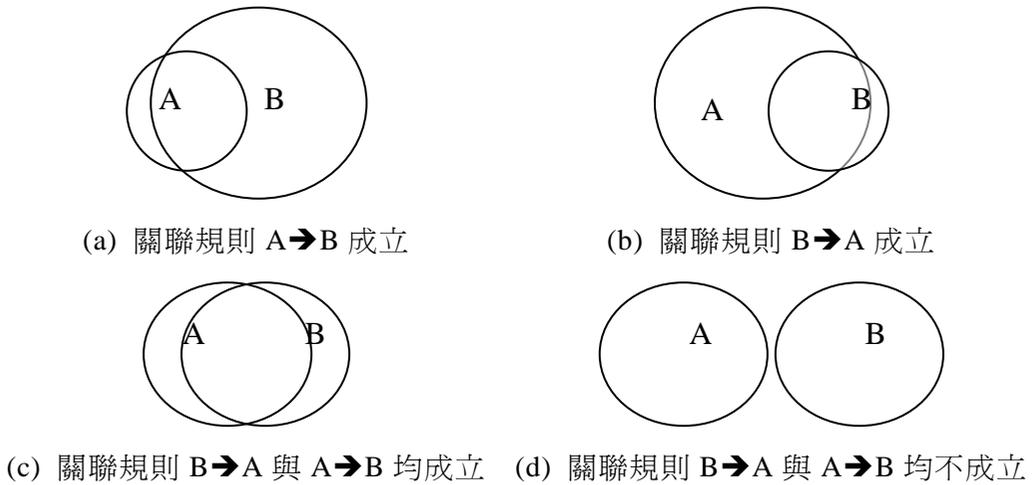


圖 4 關聯規則方向性示意圖

- (一) 由圖 4(a)可知，關聯規則 $A \rightarrow B$ 成立，也就是說當 A 出現時，B 通常也會伴隨著一起出現，但 $B \rightarrow A$ 不成立。
- (二) 由圖 4(b)可知，關聯規則 $B \rightarrow A$ 成立，也就是說當 B 出現時，A 通常也會伴隨著一起出現，但 $A \rightarrow B$ 不成立。
- (三) 由圖 4(c)可知，關聯規則 $B \rightarrow A$ 與 $A \rightarrow B$ 均成立。
- (四) 由圖 4(d)可知，關聯規則 $B \rightarrow A$ 與 $A \rightarrow B$ 均不成立。

在此以我們所發掘出一個關聯規則為例，探討關聯規則是否能預測股價漲跌。此一關聯規則為「鼎元跌 \rightarrow 嘉晶跌」，信賴度為 0.606，支持度為 20。以 2005 年的股市資料來驗證此關聯規則是否能發揮預測的功效，所得的結果如圖 5。由圖可知，當鼎元下跌時，嘉晶很可能下跌 (49/80)；如果沒有下跌，而是持平，則隔日很可能下跌 (12/19)。

觀察了前述例子，當我們在產生的結果中發現了同時漲或同時跌的關聯規則時，有時雖然無法在當天就符合同漲或同跌的關係，但在觀察其下一個交易日的表現，可以發現到可能發生「補漲」或「補跌」的情形。

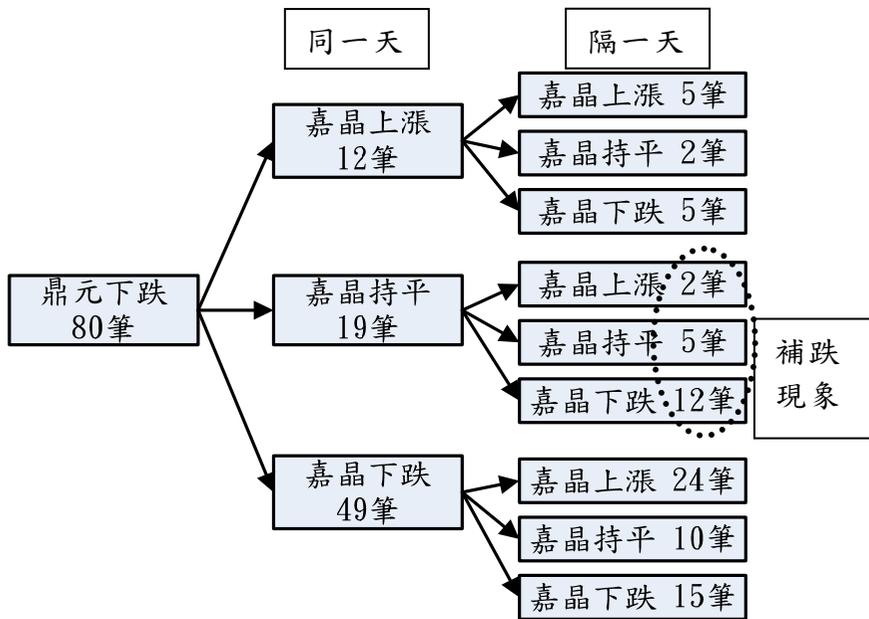


圖 5 股市補跌範例結果圖

爲了驗證此一結果的預測力是否有效，本研究從產生的關聯規則中，挑出 15 個同漲的關聯規則與 15 個同跌的關聯規則，對表 1 進行檢定分析。以下以表 1 的假說 1「依據關聯規則預測當天應該下跌（上漲），而未下跌（上漲）者，隔天會發生下跌（上漲）的機率大於上漲（下跌）的機率。」爲例說明假說的檢定過程。

以 15 個同漲的關聯規則中的「亞光 UP=>大立光 UP」爲例，其詳細的統計資料如表 6 所示。由表可知在亞光上漲的情況下，大立光當天未上漲的天數爲 26 天（=3+7+5+7+3+1），隔天大立光上漲者有 10 天（=3+7），下跌者 6 天（=5+1），故「關聯規則預測當天應該上漲，而未上漲，隔天發生上漲的機率」=10/26=0.385，「關聯規則預測當天應該上漲，而未上漲，隔天發生下跌的機率」=6/26=0.231。

接著，再以 15 個同跌的關聯規則中的「春雨 DOWN=>高興昌 DOWN」爲例，其詳細的統計資料如表 7 所示。由表可知在春雨下跌的情況下，高興昌當天未下跌的天數爲 81 天（=3+10+9+14+19+26），隔天高興昌上漲者有 17 天（=3+14），下跌者 35 天（=9+26），故「關聯規則預測當天應該下跌，而未下跌，隔天發生上漲的機率」=17/81=0.210，「關聯規則預測當天應該下跌，而未下跌，隔天發生下跌的機率」=35/81=0.432。

表 6 在亞光上漲下，大立光當天與隔天的漲跌統計表（括號中數字為天數）

		隔天		
		上漲	持平	下跌
當 天	上漲	21	17	14
	持平	3	7	5
	下跌	7	3	1

表 7 在春雨下跌下，高興昌當天與隔天的漲跌統計表（括號中數字為天數）

		隔天		
		上漲	持平	下跌
當 天	上漲	0	1	2
	持平	3	10	9
	下跌	14	19	26

表 8 假說 1 之成對母體平均數差異 t 檢定

	依據關聯規則預測當天應該下跌（上漲），而未下跌（上漲）者，隔天發生下跌（上漲）的機率	依據關聯規則預測當天應該下跌（上漲），而未下跌（上漲）者，隔天發生上漲（下跌）的機率
平均數	0.369355	0.293599
變異數	0.019077	0.00802
觀察值個數	30	
t 統計	2.250898	
P (T<=t) 單尾	0.016072	

將挑出 15 個同漲與 15 個同跌的關聯規則都計算上述二個機率，得到 30 組資料，接著作「成對母體平均數差異 t 檢定」如表 8 所示。由表可知，與規則的漲跌趨勢『相同』的機率做為隨機變數，其平均值 0.369，「相反」者 0.293，t 統計量為 2.25，單尾檢定結果為 0.0161，可以證明假說 1 是相當顯著的。

表 9 是所有假說的統計檢定結果，可知有四個假說通過信賴水準 5% 的門檻，分別是假說 1、假說 3、假說 4、假說 8。證實以關聯規則來預測股價漲跌是具有可行性的。

但上述檢定僅能證明本研究提出的規則適用在 2005 年當年度的股價表現上。有鑑於此，本研究利用 2006 年 1 至 6 月的每日收盤價格，對表 1 的假說 8 與假說 9 進行檢定，結果如表 10 所示，可知假說 8 在 2006 年上半年仍然成立，但假說 9 不成立。可見假說 8 「當關聯規則預測當天應當『下跌』的股票，在當天『持平』者，有很高的機率隔天『下跌』的機率大於『上漲』的機率。」有很高的可信度。

表 9 假說檢定

假說	顯著性	通過信賴水準 5%的門檻
1.	0.0161	*
2.	0.3270	
3.	0.00116	*
4.	0.0349	*
5.	0.2883	
6.	0.6663	
7.	0.7045	
8.	0.0335	*
9.	0.0782	

表 10 假設檢定 (2006 年 1 月~6 月資料)

	假說	顯著性
8.	依據關聯規則預測當天應該下跌，而持平者，隔天會發生下跌的機率大於上漲的機率。	0.0007
9.	依據關聯規則預測當天應該下跌，而上漲者，隔天會發生下跌的機率大於上漲的機率。	0.6825

伍、結論與建議

一、結論

(一) 關聯規則與產業關聯圖

1. 股價漲跌形成的關聯規則會顯現產業關聯。例如我們發現關聯圖中產生了化工、光學鏡頭、光碟、IC 設計、IC 測試、石英元件、液晶面板、飯店娛樂、鋼鐵、營建等產業群聚。
2. 並非所有產業均出現在產業關聯圖上，例如金融業，這有可能是此產業內各家公司的表現差異很大，不容易產生股價漲跌關聯。

3. 一些知名的大企業都未出現在產業關聯圖上，例如台積電、聯電、台塑、宏碁等，這有可能這些股本大的企業漲跌幅度較小，較少交易日的漲跌幅能達到本研究設定的漲跌定義，因此不易通過「支持度門檻」以形成頻集。未來可考慮將漲跌的定義改成以個股為基準來設定，即以使該個股被定義為「上漲」與「下跌」的交易日各達 25% 的股價變化百分比做為漲跌的定義門檻。
4. 產業關聯圖有個重要的特性就是皆為同漲或同跌的關聯，並沒有發現上漲和下跌有關聯的情形。
5. 產業關聯圖中也存在部份的巧合關聯，這可能是因為某些個股的股價恰好在同一段期間發生上漲或下跌而形成關聯規則。若能將資料收集期間拉長或是將支持度的門檻值提高，應可減少這類規則。

(二) 關聯規則之統計檢定

經由統計分析結果，顯示有四個假說過信賴水準 5% 的門檻：

假說 1：依據關聯規則預測當天應該「下跌（上漲）」，而「未下跌（上漲）」者，隔天會發生「下跌（上漲）」的機率大於「上漲（下跌）」的機率。

假說 3：依據關聯規則預測當天應該「下跌」，而「未下跌」者，隔天會發生「下跌」的機率大於上漲的機率。

假說 4：依據關聯規則預測當天應該「下跌（上漲）」，而「持平」者，隔天會發生「下跌（上漲）」的機率大於「上漲（下跌）」的機率。

假說 8：依據關聯規則預測當天應該「下跌」，而「持平」者，隔天會發生「下跌」的機率大於「上漲」的機率。

歸納這四個假說可以發現，假說 1 與假說 4 支持「補漲補跌」論述，而假說 3 與假說 8 只支持「補跌」論述，因此「補跌」獲得較「補漲」更強的支持。

(三) 關聯規則在股價漲跌之預測

本研究以 2005 年漲跌記錄產生的關聯規則對次年（2006 年）的交易日進行預測，發現假說 8 仍然通過信賴水準 5% 的門檻，證實本研究所產生的關聯規則可以預測股價的「補跌」現象，即當個股股價依關聯規則被預測為「下跌」時，有很高的機率會在當日「下跌」，若股價僅為「持平」時，其隔日有很高的機率會發生「下跌」（即補跌）的情況。

二、建議

建議後續研究方向如下：

- (一) 可探討在不同年度產生的關聯規則是否具有相似性。
- (二) 本文已初步證實已股價漲跌產生的關聯規則預測股價漲跌具有可行性。但本文只有 2005 年產生的規則預測 2006 年上半年的股價漲跌，因此未來應增加更多年度的預測，例如分別以 2001，2002，2003，2004，2005 年產生的規則，去預測 2002，2003，2004，2005，2006 年的股價漲跌，以精確評估其預測能力。
- (三) 可評估以股價漲跌產生的關聯規則做股票買賣決策依據的獲利能力。

參考文獻

一、中文部份

1. 蕭正南(1998)，資料探勘應用於股市股價趨勢之研究，輔仁大學資訊管理學研究所碩士論文。

二、英文部份

1. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases, Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 207-216.
2. Agrawal, R., & Srikant, R. (1994). Fast algorithms for Mining Association Rules, Proceedings of the 20th International Conference on Very Large Data Bases, 487-499.
3. Han, J., Cai, Y. D., & Cercone, N. (1992). Knowledge Discovery in Database An Attribute-Oriented Approach, Proceedings of the International Conference on Very Large Data Bases, Vancouver, Canada, 547-559.
4. Lu, H., Han, J., & Feng, L. (1998). Stock Movement Prediction and N-dimensional Inter-transaction Association Rules, Proceedings of ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 12:1-12:7.
5. Miller, R. J., & Yang, Y. (1997). Association Rules over Interval Data, ACM SIGMOD

Record, 26(2), 452-461.

6. Tung, A. K. H., Lu, H., Han, J., & Feng, L. (1999). Breaking the Barrier of Transaction: Mining Inter-transaction Association Rules, Proceedings of 1999 ACM SIGKDD International Conference Knowledge Discovery and Data Mining, 297-301.

2006年09月05日收稿

2006年11月17日初審

2007年01月15日複審

2007年01月18日接受