

# 第一章 緒論

## 1.1 研究動機與背景

證券市場的交易活絡與否，可以看出國家整體經濟的狀態與企業發展方向。由於低利時代的來臨，再加上通貨膨脹的影響，使得人們對於投資抱持著審慎保留的態度，也至於財富不斷地縮水造成社會經濟退化等現象。故大部份投資人多轉向多元的投資理財規劃，其中以證券投資更是為國內民眾首選的投資方式。但證券市場本身受到許多因素的影響（包含基本面、資金面、技術面、消息面及政治面等），其中不確定性很高常使得投資人蒙受很大的損失。

近年來常有學術研究在此方面進行研究並且發表成果，例如以人工智慧技術應用於股票價格漲跌預測[1]-[5]、基於技術指標分析的選股策略[6][7]、股票買賣時間點的建議[8]-[11]及股票價格漲跌轉折點預測[12]-[14]等。但這些研究大多著重於數值面的分析，甚少將消息面因素納入研究的範圍之內，其原因為當時的消息資訊之數位化尚未普遍。以證券交易市場而言，新聞事件的發生往往影響投資人的決策，而投資人的決策就會影響股票價格的波動。投資人必須廣泛蒐集各項股市新聞及消息，透過審慎的分析與判斷才能夠從中獲利。近年來網際網路的盛行，造就現今資訊發達的年代。只要是閱讀者有需要，即可透過各式各樣的傳播媒介獲得所需的資訊。因此，有關資料探勘（Data Mining）的研究逐漸受到重視，其研究的範圍廣泛觸及各個不同的層面，像是商品顧客調查、學習成果探討與人類遺傳基因等等。探討分析這些資料探勘與股市交易的研究文件中發現，有關股市消息面與數值面的資料探勘研究，少有人作這方面的探討，所以研究股市消息面與股市交易活動是一個值得研究探討的議題。因此，本研

究計畫將股市的消息面與交易活動的數值面結合，期望從中可以探勘出兩者之間所隱含的關聯規則。利用資料探勘中關聯式規則的方法來作為本研究的核心，將其資料探勘的結果提供給閱讀者一個具有精確性與可信度高的資訊，藉由消息面與數值面的隱含關聯規則來協助使用者，並且提供作為在股市買賣交易上的決策參考資訊。

## 1.2 研究目的與方法

本計劃研究的目的是在於能夠從大量的新聞事件之中找出寶貴重要的訊息，並且可以透過資料探勘方式，從這些消息面的元素裡頭找尋出與交易活動的隱含關聯性。結合資料擷取、資料儲存、資料整合分析與資料探勘等多層面技術，來完成本論文之研究系統與實驗，如圖 1-1 所示。

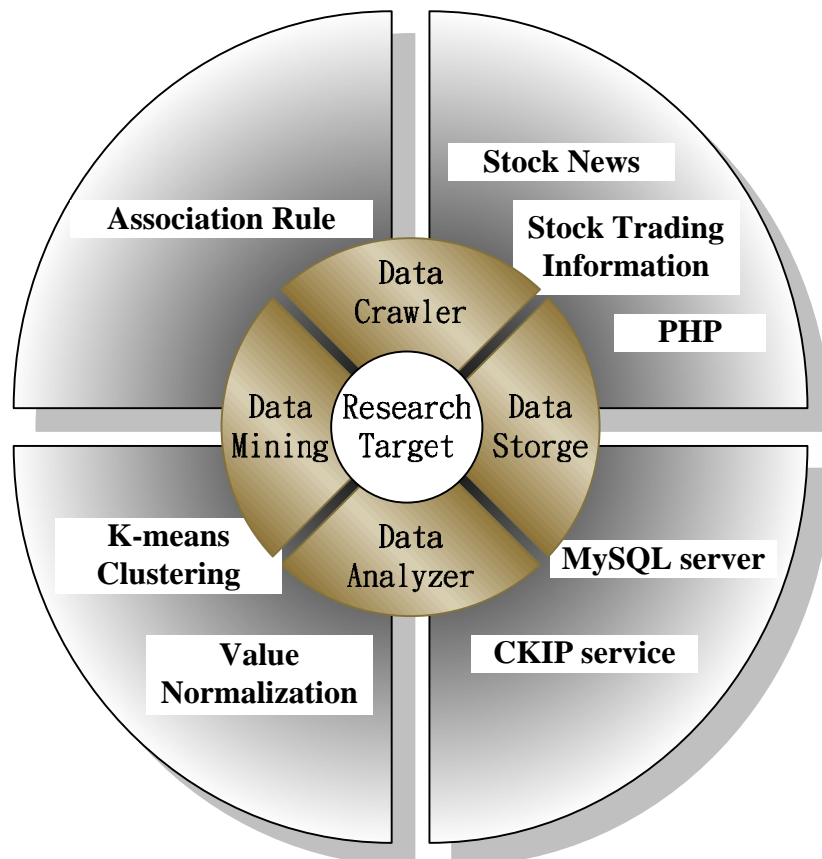


圖 1-1 理論技術架構

由於網路資訊的發達，資訊服務相對蓬勃發展，也帶動所有相關的產業技術一同成長。為了達到擷取資訊的目的，可以利用適當的工具或方法

來加以完成目標，其中目前有關撰寫網際網路程式的程式語言 PHP (PHP: Hypertext Preprocessor) 正符合本研究的需求。因為它具有跨平台的特性以及豐富的函式庫，在對於網際網路的程式撰寫上有所幫助。由網際網路上所獲得的資訊必須要有系統性的儲存，其解決的方式是將這些資訊儲存至資料庫之中，然而現今資料庫種類繁多各有不同的特性。因此，本研究採用當中與 PHP 程式語言相容性高且效能好的 MySQL server 作為本研究的資料庫。運用關聯性資料庫結構，可以大幅提昇資料查詢的速度與效能，這對本研究的系統有其重要的影響。然而這些有關股市新聞事件與交易訊息的資料量是龐大且複雜的，所以必須將所蒐集的新聞事件簡化分類。在分類之前必須先找出各則新聞事件的關鍵詞彙，這可透過中央研究院自然語言研究小組所開發的中文斷詞系統 CKIP[47]，達成找尋關鍵詞彙的此項需求目的。由於股市交易狀態有上漲也有下跌，依據本研究的範圍，必須將每日個股交易漲跌幅度正規化 (Normalization)。針對漲跌的幅度進行群聚分類 (Clustering)，把整個漲跌幅度做程度上的區分，並且針對這些群聚分類的範圍結果作為資料探勘 (Data Mining) 分析的數據。由關聯式規則運算所得的極大項目 (Large Itemsets) 與支持度 (Support Level)、信賴度 (Confidence Level) 等判斷條件，所產生的個股新聞事件與交易活動之隱含關聯規則，將可以提供閱讀者一個股市交易買賣的參考依據。

### 1.3 研究範圍與限制

由於本研究必須要有大量的個股新聞事件與個股交易資訊，才能夠提高資料探勘的正確性與精確度。所以蒐集的資訊量必須是要足夠，才能提供本實驗做資料探勘。針對此一限制，本研究之資料庫均蒐集超過半年以上的財經新聞事件與股市交易資訊，因此可以提供一個可探勘的資訊量。

由於所收集的新聞事件包含許多不同的報導來源，同則新聞事件會有報導之時間先後與持續報導的時間長短等問題。針對此一現象，將先報導

的同類型新聞事件定義在最先出現在時間序列上的時間點，以克服時間點與時間序列之間的關係。

## 1.4 研究步驟

針對本研究之流程步驟說明如下：

### Step 1：擬定研究計畫

蒐集相關文獻確定研究方向，擬定研究方法與步驟。

### Step 2：軟體模擬

1. 蒐集所需的股市交易資訊與個股新聞事件，並且儲存至系統資料庫之中。
2. 整合所蒐集的相關資訊並對漲跌幅度作正規化處理。
3. 撰寫 K-means 群聚程式，將股市交易漲跌幅度群聚分類，提供資料探勘的運算範圍。
4. 撰寫關聯式規則程式，對目標股市代碼之各個群聚的漲跌幅度範圍進行資料探勘。

### Step 3：歸納結論與建議

對於關聯式規則運算的結果進行分析與探討。

### Step 4：撰寫研究報告

將本文所設計之系統、模擬結果及實驗加以彙整並撰寫報告。

本論文的研究步驟流程如圖 1-2 所示：

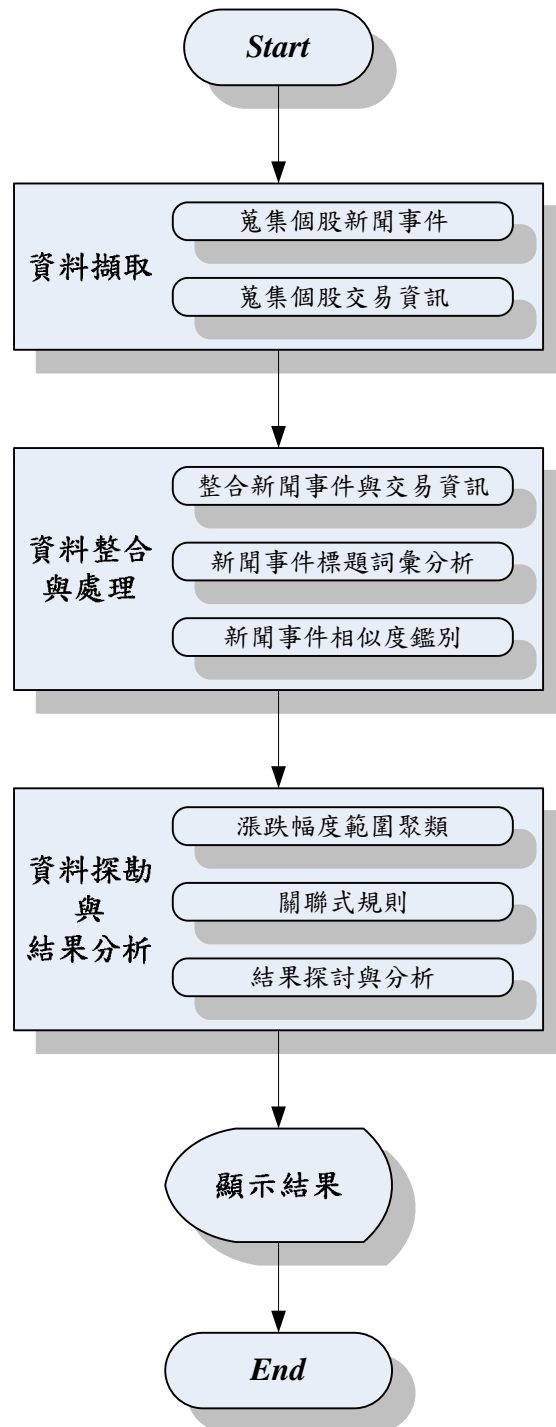


圖1-2 研究步驟流程圖