

連鎖超市導入資料探勘之研究

A study of applying data mining approach in supermarket chains

研究生：沈秀蓁

指導教授：葉惠忠

摘要

受到全球經濟不景氣的影響，許多行業的成長停滯，甚至嚴重到面臨衰退的命運。以零售業三種主要業態來看，「超級市場」長期夾擠於「量販店」與「便利商店」之間。在日趨競爭的環境下，如何利用其特有優勢、突破現有僵局找尋利基是備受矚目的。本研究以高雄地區的知名連鎖超市為研究對象，希望藉由資訊科技與管理科學的助益，導入資料探勘的技術，以強化個案超市的競爭能力。

本研究使用資料探勘之購物籃分析及預測模型，利用個案超市的交易紀錄與顧客抽樣調查，找出商品之間的關聯係數及建立認同卡的目標客戶模型，以提供相關商品組合的建議及作為認同卡推行的參考。希望藉由研究結果提供個案公司新的經營策略新的思維，強化其在未來零售市場之競爭力，同時也給予其他類以業者導入資料探勘技術實務上的參考。

關鍵字：資料庫行銷、關聯法則、資料探勘、連鎖超市

Abstract

Because the global economic is downturn, the growth of many companies is limited. Supermarket is between convenience store and hypermarket, how it should to find the advantage is very important in the more competitive environment. In this study, I use the well-known supermarket chains in Kaohsiung as the as the research object, I wish to take advantage of information technology and management science to help import data mining technology, to strengthen the competitiveness of the supermarket.

In this study, I use data mining of market basket analysis and forecast models to find the correlation coefficient between the goods and to establish the target customer model of identity card by records of the supermarket and sample surveys of customers to provide some suggestion of product bundling and as a reference for the implementation of identity cards. Finally, providing case company new manage strategy and thinking, also give other types of industry some practical reference to import data mining technology.

Keywords : database marketing, association rule, Data Mining, supermarket chains

結論與建議

第一節 研究結論

根據本研究目的，經資料分析與處理後，可得到下列幾點結論：

一、併買分析與商品組合

本研究藉由購物籃分析，共找出231條關聯規則，但礙於篇幅限制僅選擇其重點列舉，其規則分成「同類商品的購買」與「不同類商品的購買」兩種，在「同類商品的購買」此部份，主要找出【「統一果然優-蜜橘香瓣」→「統一果然優-蜂蜜草莓」】規則；而在「不同類商品的購買」，這部份又可分成「互補型商品」、「煙酒類商品」、「食品飲料型」與「趣味組合型商品」4類，在「互補型商品」中，主要找出【「卡旺休閒爐YKF-8」→「紅-卡旺瓦斯罐(3入裝)」】規則；在「煙酒類商品」中，主要找出【「登喜路(藍)淡煙6mg/條」→「美樂中罐500ml」】與【「金萬寶路淡煙(10支)」→「台灣啤酒500ml」】兩條規則；在「食品飲料型」中，主要找出【「迪士尼巧克力Q餅禮盒」→「波蜜果菜汁250gm-箱*罐」】規則；最後，在「趣味組合型商品」中，主要找出【「幫姆防嗆奶嘴(麥粉)」→「熊寶貝-花朵香補充包1890ml」】、【「3M玻璃專用接著劑30ml」→「(6入)麥格黑啤酒355ml」】、【「大道B&W特級玫瑰紅酒500ml」、「蘋果西打PET(1250cc)」】3條規則。以上的每一條規則均代表一個商品組合的建議，提供給個案超市作為行銷策略上之參考。

二、自有品牌與認同卡(會員卡)的目標客戶模型

在自有品牌目標客戶的分類上，應用一般化判別分析，找出真正目標客戶的正確率是94.29%；若採用決策樹分析，找出真正目標客戶的正確率是95.72%。雖然兩者方法均有不錯的成效，但決策樹分析仍略勝一籌，除了其正確率較高、分類誤差較小以外，決策樹分析找出的模型規則相對於一般判別分析更加清楚明瞭，非常適合實際應用在商業上。

而在認同卡(會員卡)目標客戶分類上，應用一般化判別分析，找出真正目標客戶的正確率是94.29%；若採用決策樹分析，找出真正目標客戶的正確率是100%，雖然其正確率很高，但其型II誤差(實際為不想嘗試認同卡的客戶，但卻被分類為會嘗試認同卡的客戶)也是100%，這表示決策樹分析沒有發揮任何分類的作用，因此，在認同卡(會員卡)目標客戶的分類上，採用一般化判別分析會是較適當的選擇。

第二節 管理意涵與對業者的建議

以往零售業者在進行商品組合時並未考量到消費者的意願，因此導致消費者可能有被強迫購買的想法，而造成消費者不愉快的經驗。然而，業者若能以資訊科技的方式，從龐大的交易資料庫中找出產品之間的關聯且進一步利用此關聯程度來進行商品的組合，便能兼顧消費者的意願推出其想要的產品，如此，不但能提升業者本身的銷售額，也可能會提升顧客的滿意度，形成雙贏的局面。除了商品組合以外，關聯法則也可以應用在商品擺設與動線的規劃設計上，以本研究所找出的規則【「統一果然優-蜜橘香瓣」→「統一果然優-蜂蜜草莓」】來看，雖然兩者屬於同品牌且同類的商品，但卻有可能沒有並排在一起，因此顧客有可能只看見其中一個，而忽略掉另一個，若能將兩者擺設在同一地方，則可降低客戶搜索商品的時間，增加商品能見度，進一步提升其銷售量。最後，關聯法則也可應用在折價券的行銷策略

上，藉由降低商品組合中其中一項商品的售價，增加其餘商品組合中其商品的銷售量，提升整體的銷售額。

零售業的競爭日益激烈，因此在新顧客越來越難拓展的情況下，如何留住既有顧客就變得十分重要。而在自有品牌的推廣與認同卡(會員卡)的推行皆能有助於提升顧客忠誠度的情況下，其兩者就成為超市業者未來應思考跟進的方向。不過，若業者沒有完整的推行計劃或策略就冒然燥進，則可能造成推行的失敗，因此，若業者能採用一般化判別分析或是決策樹分析所建立出的自有品牌與認同卡(會員卡)的顧客模型規則去預測顧客的態度，則能降低推行時可能的阻礙，增加推行成功的機率。

第三節 研究限制

1. 個案超市目前缺乏會員名單，所以無法以會員清單進行探勘，因此支持度(support)會受限。
2. 由於個案超市各家分店的商品編號與分類不盡相同，因此在交易資料的取得上，僅能以一間分店作為代表與實際整體的情況可能有些誤差。
3. 受限於人力、物力等資源的缺乏，無法處理該分店資料庫中所有的交易資料記錄，所以僅以半年的交易資料進行探勘，因此其規則的代表性仍有可議空間。
4. 在問卷方面，無法避免部份填答者可能有敷衍了事的態度而可能造成的研究偏差。

第四節 對後續研究的建議

1. 關聯法則可以應用在許多領域上，而本研究主要是針對個案超市，因此，未來研究可以考慮應用在其他的領域上。
2. 本研究在預測模型部份加入了傳統統計的區別分析做比較，未來也可以加入其他方法(例如：類神經網路)一起評比預測效果。
3. 本研究於自有品牌與認同卡(會員卡)的議題上仍有許多可以改進與深入探討的地方，留待未來研究繼續延續。