

應用資料探勘技術於一對一行銷系統

于還莒
財團法人資訊工業策進會
中壢市五權里三十八之一號
i760115@iii.org.tw

許清琦
國立台灣大學資訊工程學系
台北市羅斯福路四段一號
cchsu@csie.ntu.edu.tw

摘要

本論文是應用資料探勘(Data mining)技術挖掘銷售資料庫中的頻繁型樣(Frequent Pattern)、產生關聯規則(Association Rule),予以儲存,結合一對一行銷(One-to-One Marketing)系統,傳送一對一行銷訊息。

關鍵詞 : Data Mining、Frequent Pattern、Association Rule、One-to-One Marketing

一、前言

資料探勘技術在企業上常被用來挖掘經營過程中比較不明顯的趨勢,利用此發現的趨勢,企業可以採取適當的措施,以提昇企業的經營績效。

而行銷則是企業維持生存的重要活動。有別於大眾行銷之無差異化的做法,一對一行銷乃是刻意區別客戶群,對不同的客戶群進行不同的行銷活動[7,8,11]。

本論文之主要目標乃在於應用資料探勘技術從企業歷史交易記錄的大型資料庫中,挖掘感興趣的關聯規則,發現銷售商品彼此之間的關聯度,並提供一套客戶群管理的機制,使企業得以運用探勘的結果建立不同類別的客戶群,並結合一套行銷的管道,讓企業得以對不同的客戶群發送不同的行銷訊息。

透過資料探勘等相關的勘測工具對企業的營運資料進行分析,再將具高代表性的分析結果應用至市場與顧客行為,企業即能調整業務流程,針對顧客實際需求進行一對一個人化服務。

現今資訊市場上已有許多資料探勘的工具[3,6,10]可以有效得到各項所需的探勘結果,也仍普遍缺乏像下列一些重要的延伸應用機制:

一、將探勘結果以某一結構存檔,將來

商家可以隨時存取。

二、提供一個方便的機制讓商家可以結合探勘結果採取後續的行銷計畫。

本論文特別針對探勘所得的關聯規則設計了一個方便的推薦機制(Recommendation System),可以讓商家進行後續的行銷計畫。

本系統主要由三大部分構成:

- 一、資料探勘子系統
- 二、客戶群管理子系統
- 三、一對一行銷通道管理子系統

資料探勘子系統主要應用既有的演算法對商品銷售資料庫進行探勘,將得到的關聯規則予以儲存;客戶群管理子系統可以將具特定消費習慣的客戶分群管理;一對一行銷通道管理子系統則提供各種訊息發送的管道,使商家能對特定的客戶群發送專屬的一對一行銷訊息。

二、關聯規則之探勘

關聯規則就是商品銷售彼此之間的關聯性,例如「顧客買麵包時便會順便買牛奶」。

關聯規則探勘的有趣之處在於,如果商家販售的商品項目很多,商品售出的關聯性並不是顯而易見的,透過資料探勘可以發現這些隱藏的規則。

某些商品之間的關聯性比較直觀,可以生活經驗推論,例如上列麵包與牛奶的例子。這樣子的組合,通常是藉由探勘結果印證之前的推論,企業經理人可藉由精確的探勘統計數據作為決策的參考或向高階主管報告的依據。

為了說明關聯規則的應用機制及其在電子商務經營上的意義,在此將一些與關聯規則有關的用詞再回顧一下[4]。

項目集(Itemset)

就是商品項目所成的集合,一個集合所代表的就是某些特定商品的組合。

例如：{牛奶，花生}、{尿布，啤酒，小魚干}。

如果一個商家賣的商品項目有 5 樣，則其各種可能的商品項目組合共有 $C(5,1) + C(5,2) + C(5,3) + C(5,4) + C(5,5) = 2^5 - 1$ 種，相當可觀。

支持度 (Support)

某項目集(商品項目組合)在所有交易中出現的頻率。

例如：被探勘的資料庫中有 10,000 筆交易，而這 10,000 筆交易中，有 3,000 筆交易被發現同時含有 {牛奶，花生}，則 {牛奶，花生} 這個項目集的支持度為 $3,000/10,000=30\%$ 。

支持度越高，表示這種組合過去被購買的頻率越高，其受歡迎的程度具有相當的代表性。

最低支持度 (Min Support)

探勘過程中可能會發掘出各式組合的項目集，但是支持度太低(由商家根據經營型態自行認定)的項目集不具代表性，因此進行探勘的軟體通常在探勘之前可以由商家設定一個支持度的門檻值，此門檻值即為最低支持度。軟體在探勘過程中便只挑出支持度達到最低支持度的項目集輸出供商家參考，支持度未達最低支持度的項目集便直接捨去不用。

頻繁項目集 (Frequent Itemset)

達到最低支持度的項目集。

在某些應用上，頻繁項目集又被稱為頻繁型樣 (Frequent Pattern)。

關聯規則 (Association Rule)

消費行為的因果推論關係。

例如：<買尿布>就會<買啤酒>。

如果純粹只從購買的商品項目推衍其關聯性的話，可以簡寫為：尿布=>啤酒。

信心度 (Confidence)

關聯規則推論的可靠程度。

例如對於關聯規則「 $A \Rightarrow C$ 」，我們要衡量買 A 的交易中同時又買 C 的比例有多高，其計算方式就是

$(\text{買}\{A,C\}\text{的交易數})/(\text{買}\{A\}\text{的交易數})$

也可以寫為

$(\{A,C\}\text{的支持度})/(\{A\}\text{支持度})$

最低信心度 (Min Confidence)

當我們獲致一組頻繁型樣的時候，要進一步去推衍各種可能的關聯規則的組合並分

別計算其信心度，信心度太低的關聯法則便沒有代表性。

因此進行探勘的軟體通常可以由商家設定一個信心度的門檻值，此門檻值稱為最低信心度，軟體在探勘過程中便留下信心度達到最低門檻值的關聯規則供商家參考，信心度未達門檻值的關聯規則便直接捨去。

關聯度 (Correlation)

關聯規則中推論命題彼此之間的關聯性。

對於一關聯規則

$\text{LeftItemSet} \Rightarrow \text{RightItemSet}$

$$\text{Correlation} = \frac{\text{Support}(\text{LeftItemSet} \wedge \text{RightItemSet})}{(\text{Support}(\text{LeftItemSet}) * \text{Support}(\text{RightItemSet}))}$$

若 $\text{Correlation} > 1$ 則表示推論項目集彼此正向相關，若 $\text{Correlation} = 1$ 則表示推論項目集彼此獨立，若 $\text{Correlation} < 1$ 則表示推論項目集彼此反向相關。

整個探勘的過程主要分為三個階段：

1. 找出所有的頻繁項目集。
2. 根據頻繁項目集產生關聯規則，並分別計算其信心度、關聯度。
3. 儲存探勘獲得之關聯規則

關聯規則的探勘的核心技術就是頻繁項目集的探勘，這在過去數年一直有許多相關的論文發表。

IBM 的 R. Agrawal 在 1994 提出的 Apriori 演算法 [1] 是這個領域的重要時點。這個演算法使用的主要概念就是藉由不斷地掃描資料庫的交易記錄，計算出各種可能的商品項目組合(稱為候選型樣或候選項目集)的支持度，留下各個支持度達到門檻值的頻繁項目集。

SFU 的 Han 教授於 2000 年發表 FP-tree 的做法 [5]，不需不斷地掃描資料庫，也不需產生候選項目集，算是重要的突破。

台大資工所的許清琦教授與曾繁鎮博士生於 2001 年又改進了一個更簡潔的 FP-list 的方法 [12]，使其探勘的效能更好，本論文係應用此方法進行探勘。

每次探勘結果可能會得到許多頻繁型樣，採取 XML 的格式 [2,9,13,14,15] 儲存，其 Document Type Definition (DTD) 如下：

```
<!DOCTYPE FPL [  
<!ELEMENT FPL (FP, TotalCount, Threshold)>  
<!ELEMENT FPL (SKU)+>
```

```

<!ALLIST FPL Count (CDATA) #REQUIRED>
<!ELEMENT SKU (#PCDATA)>
<!ELEMENT TotalCount (#PCDATA)>
<!ELEMENT Threshold (#PCDATA)>
]>

```

每一組頻繁型樣存入一組 FP 標籤中，SKU 存頻繁型樣中商品項目的料號，屬性 Count 表此頻繁型樣的支持度，例如

```

<FP Count="507">
  <SKU>108</SKU>
  <SKU>115</SKU>
</FP>

```

表示此組頻繁型樣中含有兩項商品項目，其料號分別為 108 及 115，支持度為 507，就是在 507 筆交易中含有這個型樣。

XML 中並另有兩個標籤存放探勘的總交易筆數及探勘時所設定的支持度門檻值。例如：

```

<TotalCount>6515</TotalCount>
<Threshold>500</Threshold>

```

表示探勘的總交易筆數為 6515 筆，探勘者所設定的門檻為 500。

系統並定義一個 XSLT[16]用來將頻繁型樣按其支持度遞減排序、同一組頻繁型樣內按料號遞增排序輸出：

```

<?xml version="1.0"?>
<xsl:stylesheet
xmlns:xsl="http://www.w3.org/TR/WD-xsl">
<xsl:template match="/">
<FPL>
<xsl:for-each select="//FP" order-by = "number(@Count)">
<FP>
<xsl:attribute name="Count">
<xsl:value-of select="@Count"/>
</xsl:attribute>
<xsl:for-each select="SKU"
order-by="+ ./text()">
<SKU><xsl:value-of/></SKU>
</xsl:for-each>
</FP>

```

```

</xsl:for-each>
<TotalCount>
<xsl:value-of select="//TotalCount"/>
</TotalCount>
<Threshold>
<xsl:value-of select="//Threshold"/>
</Threshold>
</FPL>
</xsl:template>
</xsl:stylesheet>

```

以下是某一交易資料庫所得的結果

```

<FPL>
  <FP Count="1468">
    <SKU>115</SKU>
  </FP>
  <FP Count="1111">
    <SKU>109</SKU>
  </FP>
  <FP Count="1013">
    <SKU>108</SKU>
  </FP>
  <FP Count="832">
    <SKU>118</SKU>
  </FP>
  <FP Count="507">
    <SKU>108</SKU>
    <SKU>115</SKU>
  </FP>
</TotalCount>6515</TotalCount>
<Threshold>500</Threshold>
</FPL>

```

根據前面所得到的各組頻繁型樣，求出各組頻繁型樣所可能衍生的關聯規則，將各個規則存入一資料表中，資料表中含下列資料項。

- 探勘序號(用以區分各次探勘)
- 此規則所對應的頻繁型樣(XML)
- 支持度次數

- 支持度百分比
- 信心度百分比
- 關聯度百分比

將關聯規則存檔後即可進行應用。

三、關聯規則之應用

對於一條關聯規則的取捨，有三個重要的參數需要考慮：

- 1.支持度：支持度夠大，表示這條規則具有相當程度的代表性。
- 2.信心度：信心度夠大，表示此規則的因果關係夠強。
- 3.關聯度：關聯度要大於 1，表示是正向相關。

這三項參數都存在本系統的探勘結果中。只要將它們顯示出來，並提供各種排序方式，商家便可以很方便地挑出自認為有意義的規則做各項應用。

像「A=>C」這樣的關聯規則，表示「買 A 的也會同時買 C」，這時商家可以有列措施可以考慮：

一、商家可以考慮將兩樣商品綁在一起促銷，這種促銷方案可以提高消費者的購買意願。

二、商家可以找出過去交易中，曾經買過 A 但尚未買 C 的消費群，直接對他們來促銷 C，這樣促銷成功的機會較大。

本系統提供一個介面，可以讓商家在檢視有興趣的關聯規則時進一步將與此關聯規則有關的客戶群定義出來，以作為未來進行促銷活動的對象，這樣的客戶群稱為消費群。

本系統建立兩個資料表來存放消費群的定義。

本系統另外設計了一個預存儲程序可以在需要的時候，執行此預存程序到客戶資料庫中擷取所定義的客戶群。

四、一對一行銷系統

當有了促銷對象，便可以規劃行銷計畫，並透過各種管道進行一對一行銷。

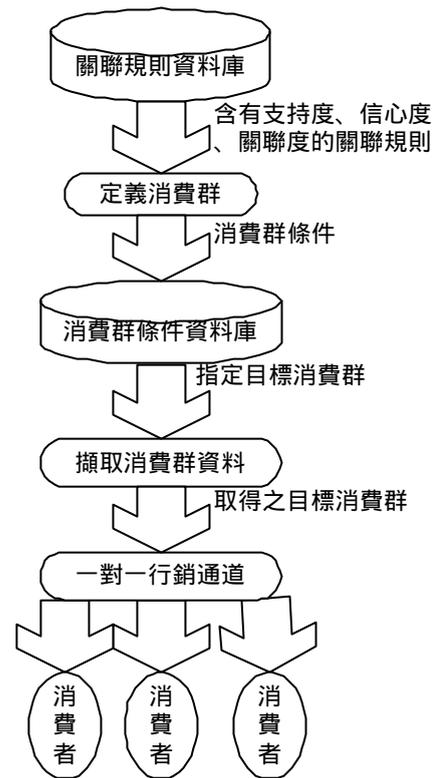
本系統目前提供兩種行銷通道可茲應用：

- 一、一對一好康報報

二、一對一電子報

前者可以將促銷的訊息以一對一的方式存成 XML 的格式，結合電子商務的前台系統，消費者登入網站時便可以看到個人專屬的促銷訊息。

後者則是將促銷訊息以一對一客製化的方式主動以電子郵件發送給各個消費者，消費者收到的電子報訊息是專為個人量身訂做的資訊，不會因為資訊太多而影響消費者閱讀的意願。



<圖 1>關聯規則之應用機制

<圖 1>顯示本系統可以將關聯規則定義的消費群擷取出來，結合一對一行銷系統傳送一對一的行銷訊息。

模組化的結果，可以將客戶群的定義及一對一行銷通道做出區隔，以利系統之延展性。

五、客戶群管理

區隔客戶群可以說是一對一行銷的關鍵動作。區隔客戶群的核心思想在於每一個客戶的個人特質不同、喜好不同、對商家的貢獻度也不同，不應同等對待。

因此區隔客戶群可以用各種不同的標準，可以根據客戶的重要性、貢獻度、潛在消費能力、過去的消費記錄、客戶本身的特質等來做區隔。

本系統原先的構想僅係要設計一個方便彈性的機制將資料探勘所得的關聯規則予以轉化成可以進行商務促銷活動。然而多變的企業經營環境所需的是更具彈性的客戶群分類機制，因此本系統目前提供下列幾種方式以供商家區隔客戶群：

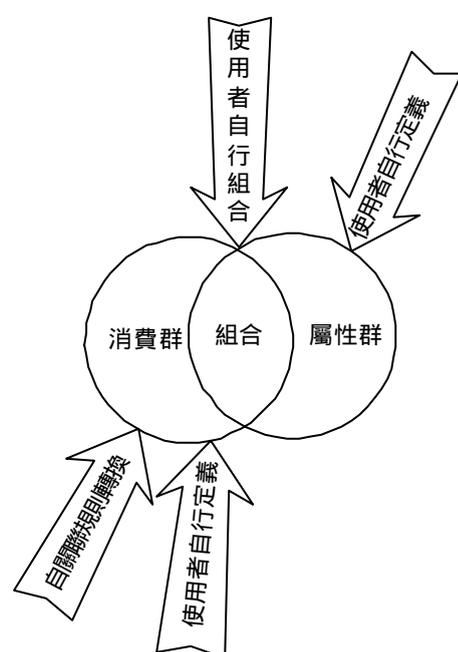
一、根據客戶本身的特質作區分，稱為屬性群

二、根據客戶過去的消費記錄作區分，稱為消費群。消費群又分正向表列消費群(也就是曾經買過哪些商品)及負向表列消費群(也就是不曾買過哪些商品)。

三、用戶自行挑選客戶組成客戶群。

四、用戶自行以 SQL 指令定義客戶群(給具 SQL 知識的用戶)。

五、用戶根據上列各項條件組合成自己想要的複雜條件的客戶群，例如具有哪些屬性且曾購買過哪些商品的客戶。



〈圖 2〉關聯規則之應用機制

本系統最重要的特色就是可將資料探勘的結果輕易轉換為客戶群的定義，因此屬性群的定義可以由商家依自己的認定來設定目標客戶群的條件，也可由多維關聯規則的探勘的結果由系統自動轉換，由商家決定是否定義成

屬性群。

消費群的定義也可以由商家自行設定條件或在關聯規則產生時由系統向商家推薦變成消費群的定義儲存。

六、結論與未來工作

一對一行銷有別於傳統的大眾行銷，在企業電子化的過程中又可以結合各種不同的電子化行銷通道，以建立一對一行銷系統。

本系統應用資料探勘技術，挖掘商品交易的關聯規則後予以儲存，再將關聯規則轉換為消費群定義用以區分消費群，並可結合一對一好康報報及一對一電子報等行銷通道對客戶進行一對一行銷。

本系統未來預計將朝下列方向繼續增進。

一、行銷活動回饋分析機制

行銷活動的效益通常是反應在營業額的提昇上，然而只是看整個營業額未免過於籠統，而一對一行銷最重要的理念就是找對的人賣對的商品或服務，因此要如何持續評估是否將對的訊息傳遞給了對的人，便是一件重要的課題，這將是本系統後續的努力方向。

二、客戶價值權重評量模型

一對一行銷的重要精神之一就是讓客戶覺得這個行銷活動好像是專門為自己量身訂做的；從商家的角度看，一對一行銷也就是要營造客戶的最高價值，長久保有高價值的客戶。然而客戶的價值要怎樣衡量呢？因此研製一個客戶價值量化的模型便成為客戶價值管理的重要課題，這也是本系統後續有興趣的主題。

三、結合其他資料探勘方法

資料探勘領域尚有許多有趣的方法，而行銷系統的延伸則為完整的顧客關係管理系統，一個永續經營的企業應該對其客戶及相關的銷售型樣永遠感興趣，而客戶區隔及分析正可配合其他的資料探勘的方法來達成。

七、參考文獻

- [1] R. Agrawal and R. Srikant. "Fast Algorithms for Mining Association Rules in Large Databases" Proc. of the 20th Int'l Conference on Very large Databases, pp.487-499, Santiago, Chile, Sep 1994
- [2] Patrick van Amstel, Pim van der Eijk, Evert

- Haasdijk, David Kuilman. "An interchange format for cross-media personalized publishing"
- [3] J. Han. "DBMiner"
<http://db.cs.sfu.ca/DBMiner>
- [4] J. Han and M. Kamber "Data Mining Concepts and Techniques" Morgan Kaufmann Publisher, pp.229-230, 2000
- [5] J. Han, J. Pei, and Y. Yin. "Mining Frequent Patterns without Candidate Generation" Proc. of ACM SIGMOD, pp.1-12, Dallas, TX, May 2000
- [6] IBM. "Intelligent Miner"
<http://www-4.ibm.com/software/data/iminer>
- [7] Jupiter Researcher and Advisory Services.
<http://www.jup.com>
- [8] Graeme McCorkell "Direct and Database Marketing", Chinese Version, pp.87-88, 2000
- [9] Microsoft Official Curriculum. "Building XML-Based Web Application"
- [10] NCR. "Teradata", <http://www.teradata.com>
- [11] Amrit Tiwana. "The Essential Guide to Knowledge Management: E-business & CRM applications", Prentice-Hall, pp.23-34, 2001
- [12] Fan-Chen Tseng and Ching-Chi Hsu. "Creating Frequent Patterns with the Frequent Pattern List". Proc of Asia Pacific Conference of Data Mining and Knowledge Discovery, pp.376-386, Hong Kong, 2001
- [13] US Navy. "MIL-PRF-87269"
<http://navycals.dt.navy.mil/cals/calsstds.html>
- [14] World Wide Web Consortium. "Document Object Model (DOM) "
<http://www.w3.org/DOM>
- [15] World Wide Web Consortium. "Extensible Markup Language (XML)"
<http://www.w3.org/XML>
- [16] World Wide Web Consortium. "XSL Transformations (XSLT)"
<http://www.w3.org/TR/xslt.html>