

植基於類神經網路之顧客交易特徵分析機制

呂慈純

朝陽科技大學資訊管理研究所
tclu@cyut.edu.tw

吳昆益

朝陽科技大學資訊管理研究所
s9354610@cyut.edu.tw

摘要

顧客分群對於企業的目標行銷或市場區隔是一個重要的關鍵步驟，雖然目前已有許多相當不錯的分群方法被提出來，但是這些方法大都著重在數值資料的運算上，而不是由商業目標做為考量。本研究提出了一個改良的分群方法 TPCSNN，這個方法是植基於一個以顧客過去的交易特徵為基礎的分群方法，先過濾出足以代表此客戶交易特徵的交易資料，再利用降低群間相關係數與提高群內密度的方式，反覆計算來達到客戶分群。接著進一步運用類神經網路的技術挖掘出的消費行為特徵，所得之結果可供分群新客戶之用。讓使用者不用再花大量的時間成本，即可快速進行新顧客分群動作。除此之外，本研究也針對分群的結果加以分析與解釋。

關鍵字：市場區隔、顧客區隔、關聯規則探勘、類神經網路

Abstract

Customer segmentation is a key element for target marketing or market segmentation. Although there have been quite a lot of ways available for segmentation today, most of them emphasize numeric calculation instead of commercial goals. In this study, we propose an improved segmentation method called TPCSNN. This is a segmentation method based on customer's historical transaction patterns. First of all, it filters transaction data for records with typical patterns. Next, it reduces inter-group correlation coefficient and increases inner cluster density to achieve customer segmentation by iterative calculation. Then, it utilizes neural network to dig patterns of consumptive behaviors. The results can be used to in the segmentation of new customers. By this way, customer

segmentation can be implemented in very short time and costs little. Furthermore, the results of segmentation are also analyzed and explained in this study.

Keywords: market segmentation, customer segmentation, association rule mining, neural network

1.前言

消費市場是一個沒有固定邏輯且快速變動的環境，在大部分的情形下，市場上買家的需求形形色色，若只單純提供一、二項產品，勢必無法滿足所有消費者的需求，但過多的商品種類又會增加公司營運的負擔與風險。因此為了同時滿足不同客戶的需求及考慮公司本身的能力，必須對消費者市場做“區隔”的動作，將市場切割成數個分群，針對不同的族群需求擬定適合的行銷策略。

Kotler 所提出的策略行銷的核心就是 STP 三個步驟-亦即區隔(Segmentation)、目標(Targeting)、及定位(Positioning)[6][8]，根據顧客的某些差異性，將現實中複雜的市場區分為數個具同質性的小市場，並從中找出企業的目標市場以及企業在此市場中的定位，這樣的策略即使到現在也依然是金科玉律。Kotler 於近年提出的新品牌行銷模式-CCDVTP(Create Communicate Deliver Value Target Profit)，創造(Create)新的溝通(Communicate)方式，傳遞(Deliver)品牌價值(Value)，然後再針對目標市場(Target)行銷，進而獲利(Profit)[7]。因此要訂定好的行銷策略，首先還是必須做好市場區隔。

現今雖然有很多市場切割的方法，但大多是建構在現有的分群演算法上，例如 K-means、DBSCAN(Density-Based Spatial Clustering of Applications with Noise)[5]等，使用者必須先針對欲解決的目標或是資料庫的特性，選擇一合適的群聚方法，選完後又會面

臨參數選擇的問題，使用者必須對問題或是資料特性十分熟悉才能找到合適的參數，進而得到最佳的結果[1]。然而這樣高的技術門檻對於一般企業幾乎是遙不可及，因此本研究將提出一個容易為一般人所理解且分群快速正確的方法，讓理論可以協助企業的經營。此外，現行演算法在設計時並不是以商業目標為考量，因此企業在應用時必須針對資料做一些調整，如此可能會導致與企業想解決的問題發生差異。所以我們針對消費市場的顧客分群問題，尋找目前已經提出的理論，希望能同時兼顧行銷與商業目的，最後找到由張簡尚偉教授與郭盛陽同學於 2004 年所提出的客戶分群演算法 TPCS(Transaction Pattern based Customer Segmentation)[3]解決了此項問題。他們利用客戶歷史交易資料，將相似交易特徵的顧客群聚在一起，且能同時考慮到行銷與商業目的。然而，他們的方法無法對分群後的結果進行分析與說明，因此本論文改良 TPCS 演算法，在原本的方法上加入交易資料的萃取機制，並對分群結果進行分析。再者，進一步應用類神經網路技術，讓使用者可以快速得到新顧客的分群結果。

本論文的架構大致如下所述：第二節為文獻探討，說明 TPCS 演算法與其他相關的技術；第三節說明所提出之改良分群方法；第四節利用模擬與實際資料評估改良方法的效果，並以實際系統來做驗證。最後第五節為結論。

2. 相關文獻

2.1 植基於交易特徵的顧客分群方法(TPCS)

TPCS 從客戶交易資料中，找出消費行為的特徵，在分群時以客戶在各群的比重，調整成以客戶為主的分群，如此可以解決單一客戶，但有不同交易資料落在不同分群的狀況。接著以消費行為特徵與相關係數(Correlation Coefficient)[2] 製作出分群關聯矩陣，衡量所有分群間的交集程度，再以改善分群關係的策略調整，使得群間關係降到最低，得到群間關係最低的顧客分群。接著，加入衡量群內關係的密度指標，透過降低群間關聯度與提高群內密度，以合併和分離雙向策略，達成顧客分群步驟[3]。這個方法有二點限制：

1. 探勘交易規則時只關注購買的商品項目而不是購買的數目。
2. 一筆交易是指該次購買的商品項目集合，與

商品購買順序無關，且包含購買單一商品的情況。

我們以 4 個顧客 4 種品號，以及 7 筆交易資料來說明 TPCS 分群的步驟：表 1(a)為原始的交易資料，Item 為每一位顧客購買的商品，A、B、C、D 為商品代號，TID 為交易代號。我們必須先把它轉換為表 1(b)的表示方法，即以 1 表示有買，0 表示沒有購買，此時加入最小支持度的判斷，將出現次數過少的品號剔除不納入計算。

表 1: 顧客交易資料表示的轉換
(a) 交易原始資料

TID	CustomerID	Item
1	甲	A、B、D
2	乙	A、B
3	丙	C、D
4	丁	B、C、D
5	甲	A、B、C
6	乙	A、B、D
7	丁	C、D

(b) 轉換為研究用的表示法

Item TID	A	B	C	D
1	1	1	0	1
2	1	1	0	0
3	0	0	1	1
4	0	1	1	1
5	1	1	1	0
6	1	1	0	1
7	0	0	1	1
Count	4	5	4	5

假設客戶初步分成了 X、Y 兩群，如表 2 所示，我們利用期望值與規則發生的機率，加上 4 個品號購買規則的排列組合，可以計算出這 2 組分群間的顧客關聯矩陣(Cluster Correlation Matrix, CCM)[3]，如表 3 所示。此外假設每個產品的利潤都為 1，表中 $P(\text{rule}_i^X)$ 為規則 rule_i 在 X 分群出現的機率， $EV(\text{rule}_i^X)$ 為期望值。我們以第 12 項規則來說明，第 12 個規則為 1100，表示有購買 A 與 B 的交易在 X 分群中的發生機率為 1/7，期望值為 2/7。

表 2: 顧客初步分群的結果

(a) cluster X

Item TID	A	B	C	D
1	1	1	0	1
2	1	1	0	0
3	1	1	1	0
4	1	1	0	1

(b) cluster Y

Item TID	A	B	C	D
1	0	0	1	1
2	0	1	1	1
3	0	0	1	1

表 3: 顧客關聯矩陣

		cluster X		cluster Y	
i	rule _i	P(rule _i ^x)	EV(rule _i ^x)	P(rule _i ^y)	EV(rule _i ^y)
1	0001	0	0	0	0
2	0010	0	0	0	0
3	0011	0	0	2/7	4/7
4	0100	0	0	0	0
5	0101	0	0	0	0
6	0110	0	0	0	0
7	0111	0	0	1/7	3/7
8	1000	0	0	0	0
9	1001	0	0	0	0
10	1010	0	0	0	0
11	1011	0	0	0	0
12	1100	1/7	2/7	0	0
13	1101	2/7	6/7	0	0
14	1110	1/7	3/7	0	0
15	1111	0	0	0	0
Correlation Coefficient (CC)				-0.1721	

接著計算群內密度，將該群資料筆數除以群內各筆交易資料與質心的距離加總，做為表達群內緊密程度的密度指標。他們是使用歐幾里得距離(Euclidean Distance)做為計算距離方式。我們可以由表 2 計算出每群的質心，其結果如表 4。

表 4: 各分群各品號的質心

(a) cluster X

ITEM	count	mean
A	4	1
B	4	1
C	1	1/4
D	2	1/2

(b) cluster Y

ITEM	count	mean
A	0	0
B	1	1/3
C	3	1
D	3	1

由表 4 可算出 cluster X 的密度為 3.0237, cluster Y 的密度為 3.6742, 2 個分群的效果為 38.9187。該演算法會把所有的分群組合兩兩配對計算，並找出分群效果最差的一組做調整，如果全部的組合都調整過，結果都不比原本的分群效果好時，整個分群就結束。

2.2 倒傳遞網路(Back-Propagation Network, BPN)

雖然 TPCS 能夠有效的進行分群，但卻無法解釋分群結果，因此我們採用類神經網路中的 BPN 進行分群的分析與學習。倒傳遞網路是屬於監督式學習網路的一種，為目前最被廣泛使用的一種網路，在診斷、預測、分類、最佳化上都有相當良好的效果。倒傳遞網路包含 3 個層：輸入層、隱藏層、輸出層，每一個層都包含數個處理單元。輸入層用來輸入外界傳入的訊息，輸出層則是將結果傳回給外界，中間可能有數個隱藏層，用來表現輸入處理單元間的交互影響。目前 BPN 是應用最普遍的網路，成功案例多，且具有學習精準度高與回想速度快的優點[1]。

3. 研究方法

本研究提出一個改良方法 TPCSNN (Transaction Pattern based Customer Segmentation with Neural Network)，希望能達到以下的目標：

1. 群間關係最小化：利用提出的顧客關聯矩陣

(CCM) 計算關聯程度，並透過降低群與群之間的相關係數，來達到分群的目的。

2. 群內關係最大化：利用歐幾里得距離產生密度指標，用以改善群內結構鬆散的問題。
3. 快速回應的系統：一般的演算法計算分群時都必須耗費大量的時間成本，本研究應用了類神經網路的技術，可以快速的判斷新顧客是屬於哪一個分群。

本方法的主要步驟如下：

1. 顧客交易資料萃取。
2. 將顧客交易資料做初步分群：將具有相同交易特徵的交易資料歸為同一群，作為初始分群。
3. 調整成為以顧客為主的分群：步驟 2 的初步分群是以交易特徵的相似來分群，因此會有同一個顧客的交易資料散在不同群的狀況，因此必須再轉換成為以顧客為主的分群，確保一個顧客只屬於一個分群。
4. 針對各分群做交易規則探勘：從每一個分群中挖掘該群顧客的交易規則特徵，以這些規則表達該群顧客的消費特性。
5. 計算群間關聯度：利用分群間的相關係數，製作分群關聯矩陣。
6. 計算群內密度：利用該分群交易資料筆數，除以該群質心與群內各筆交易資料的距離總和，做為評估群內緊密程度的密度指標。
7. 評估分群的效果：利用分群評估方式，以各分群兩兩配對的方式來衡量分群的效果是否有改善，並且挑選最差的分群配對作為改善分群結果的候選改善分群組合。
8. 改善分群效果：透過三個調整分群的策略，找出最合適的改善方式。如果經過三個調整策略都無法改善整體分群的結果，就回到步驟 7 找尋次差的分群配對組合來改善，如果都沒有改善，就一直重複到所有配對組合都嘗試過為止；如果有改善的話，就回到步驟 3 再重複計算一次。
9. 分群結果分析：提供一介面，讓使用者可以針對群與群之間，或是同一群內顧客之間的交易狀況去做分析。
10. 建立網路模型：利用前面步驟所產生出來的分群結果為訓練資料，訓練出可以快速分群的倒傳遞類神經網路。

3.1 顧客交易資料萃取

以中小企業為例，一般企業銷售的品號都相當多，交易的客戶數目也是成千上萬。因此

我們並不拿所有的品號來做分析，而是專注在被購買次數為前 N 名的主要產品，以及針對會購買這些產品的客戶來做分群。如此企業可以針對會購買熱門商品的客戶，做更進一步的分類與行銷。

另外一般 ERP 客戶在導入時，針對一些服務或費用，為了作業上的方便，也會建立品號，如安裝配送費用、贈品、租金等等。因為這樣的品號並非顧客主動購買，而是被廠商附加上去的，因此在篩選資料時也必須將這類的品號剔除。此外因為品號在編碼上，不同的分類會以不同的代號做為開頭，我們也提供此一篩選條件讓使用者選擇。資料萃取的條件如圖 1 所示。

圖 1:ERP 資料篩選程式畫面

3.2 將顧客交易資料做初步分群

在這個階段我們先將顧客的交易資料，轉換為系統所需的表示方法，並把有相同規則的交易資料合併在同一群，此時並不考慮該筆交易資料是屬於哪一個客戶。轉換方式如表 1 所述。

3.3 調整成為以顧客為主的分群

前一階段可能一個顧客會被分到不同群裡，這與我們的目的不同，為了確保顧客資料落在同一群，我們以各群內交易筆數的比例，決定哪一群是主要的分群。如果最高同百分比的有 2 群以上的狀況，則需再考慮獲利的部分，目前是以其第一群為主要分群。選擇方式如表 5，其中，顧客甲的交易資料分別歸類到分群 1、2、3 和 4 的機率為 50%，25%，12.5% 和 12.5%。因此甲的分群結果為分群 1。

表 5:選擇主要分群

顧客編號	分群	所含交易編號	百分比	主要分群
甲	1	3, 4, 5, 6	50%	←
	2	2, 3	25%	
	3	1	12.5%	
	4	5	12.5%	

3.4 針對各分群做交易規則探勘

假設所有客戶交易資料 $T=\{T_1, T_2, \dots, T_n\}$ ，每一筆交易紀錄購買的產品為 $I=\{I_1, I_2, \dots, I_m\}$ ，其中 $I_i \in \{0,1\}, i=1 \sim m$ ， m 表示產品數。當 $I_i=1$ 表示這筆交易有購買第 i 項產品，當 $I_i=0$ 表示這筆交易中沒有購買第 i 項產品，例如，假設現有提供 A、B、C、D 四項產品，如果有一筆交易資料表示式為 1101，則表示在該筆交易中有購買了 A、B 與 D 項產品。

在將所有交易資料轉換後，我們再針對每一項產品做直向加總，如表 1(b)所示，此時可再加入最小支持度 (Minimum Support) 的判斷，如果最小支持度是 0.6 且一共為 7 筆交易，則表示單項產品必須出現超過 4.2 次才會納入計算，因此表 1(b)中 A 與 C 品號就不會納入計算。在去除出現次數不高品號後，以剩下品號找出購買的交易規則。

本研究在顧客資料萃取階段，就先讓使用者篩選出銷售前幾名的品號來做分析，因為是使用者自行判斷，因此並不需要再加上最小支持度的判斷。

3.5 計算群間關聯度

相關係數是用來表現兩個變數間相互的直線關係，並判斷其緊密程度的統計方法，當其絕對值越接近 1 時，表示兩個分群間的關係越密切，為了達到群間關聯度低的目標，相關係數是越低越好。我們利用顧客每個交易特徵的期望獲利值來計算兩群間的相關係數，並用分群關聯矩陣來呈現。我們將 X 和 Y 兩個分群的相關係數用 $\rho_{X,Y}$ 來表示，計算公式如公式 (1) 所示。在公式 (1) 中 m 是指用來分析的產品數量； $E(rule_i^X)$ 和 $E(rule_i^Y)$ 表示各別交易特徵 i 的期望獲利值； μ_{rule_X} 和 μ_{rule_Y} 表示各別的交易特徵 i 的期望獲利平均值。

$$\rho_{X,Y} = \frac{\sum_{i=1}^{2^m-1} \frac{1}{2^{m-1}} (E(rule_i^X) - \mu_{rule_X}) (E(rule_i^Y) - \mu_{rule_Y})}{\sqrt{\sum_{i=1}^{2^m-1} \frac{1}{2^{m-1}} (E(rule_i^X) - \mu_{rule_X})^2} \sqrt{\sum_{i=1}^{2^m-1} \frac{1}{2^{m-1}} (E(rule_i^Y) - \mu_{rule_Y})^2}} \quad (1)$$

為了計算交易規則的期望獲利值，我們在公式 (2) 中定義了交易規則發生機率，其中 $Amount(rule_i^X)$ 表示符合該規則的實際購買商品數量， $Num(D)$ 是表示在交易資料庫中的總資料筆數，所以 $P(rule_i^X)$ 表示 X 群中第 i 種交易特徵的交易機率，如表 3 中 X 群的交易特徵 1101 出現的交易次數為 2，全部的交易筆數為 7，所以在 X 群中此交易規則發生的機率為 0.2857。接著我們用公式 (3) 來計算此交易規則的利潤，在此我們重點放在產品的購買規則上，因此暫不考慮其利潤，所以先假設各產品利潤 (Profit) 均為 1。其中 $Bit_l(rule_i^X)$ 表示在 X 集合中的規則 i 的第 l 項產品是否有購買，是用 0 或 1 表示。

$$P(rule_i^X) = \frac{Amount(rule_i^X)}{Num(D)} \quad (2)$$

$$E(rule_i^X) = \sum_{l=1}^m profit_l \times Bit_l(rule_i^X) \times P(rule_i^X) \quad (3)$$

以表 3 中的交易特徵 1101 來說，其期望的獲利值就是 $(1*1*0.2857) + (1*1*0.2857) + (1*0*0.2857) + (1*1*0.2857) = 0.8571$ 。而 X 群

的期望獲利平均值 μ_{rule_X} ，則是所有的期望獲利值 $E(rule_i^X)$ 相加，再除以所有規則數。以表 3 的 X 群來說，其期望獲利平均值就是 $(0.2857+0.8571+0.4286)/15=0.1048$ 。

3.6 計算群內密度

因顧客交易資料一般均呈現產品品號多，但交易品號少的狀況 (例如分析 50 個品號，但每一筆交易中可能不到 10 個品號是有購買的)，因此將密度的公式定為單位距離內的物件個數，如公式 (4) 所示。這個公式是利用歐幾里得距離來算出群內各筆交易與質心的總距離，如公式 (5) 所示，距離越小表示該群內成員越相似。

$$D = \frac{n}{\sum Distance} \quad (4)$$

$$其中 Distance(x_i, \bar{X}_i) = \sqrt{\sum_{j=1}^n (x_{ij} - \bar{X}_i)^2} \quad (5)$$

x_i 表示該群每一筆交易中的第 i 個品號，
 $\overline{X_i}$ 表該群所有交易中，第 i 個品號的平均值，
 n 表示該群所有的交易筆數。

3.7 評估分群的效果

我們將前面步驟所計算出的群間關聯度指標與群內密度指標合併，分子是兩群間的密度加總，其值是越大越好；分母則是兩群間的關聯程度，也就是相關係數，其值是越小越好。我們可以產生出整體評估分群效果的評估式，如公式(6)所述，其值是越大越好。

$$TCE = \frac{2}{c(c-1)} \sum_{a=1}^{c-1} \sum_{b=a+1}^c \frac{D_a + D_b}{CC_{ab}}, \quad (6)$$

TCE (Total Clustering Effectiveness)表示分群效果，公式中共有 c 群， a 、 b 是指 c 群中的其中任兩群， D_a 為第 a 群的群內密度， CC_{ab} 為第 a 和 b 群的群間關聯度。

3.8 改善分群效果

當做完一次分群之後，為了改善分群的效果，必須繼續測試目前的分群組合是否有改善的可能性。為此我們要挑出分群結果較差的組合來調整，也就是各群配對的評估值 ECE (Each Cluster Effectiveness)最差的組合。評估值如公式(7)所述。

$$ECE = \frac{D_a + D_b}{CC_{ab}} \quad (7)$$

找出最差的組合後，接著利用三種調整分群的策略來調整，看是否能改善分群的結果。如果調整完 TCE 值有改善，則接受此調整結果，再重新分配分群。如果調整完的 TCE 值都沒改善，則維持原有分群結果，再找次差的分群組合來調整。如果所有分群組合測試調整後都無法改善 TCE 的值，則我們認定此分群結果已是最佳的結果。改善分群的策略有三種，敘述如下：

1. 合併：可能原本兩群太相似，根本就該屬於同一群。拆成兩群會造成關係係數過大而降低 TCE 值。
2. 拆成三群：在兩群之中，可能交集的部分過大，導致關係係數過大，但有可能交集的部分本身就該另外成一群，因此將交集部分的顧客再獨立成一群來計算。
3. 兩群交集部分依距離區分屬於哪一群：還有一種可能就是兩群交集的部分被歸錯群，例如原本該歸到 a 群卻被歸到 b 群，因

此將交集部分以客戶為單位，看該客戶的交易資料比較接近哪一群的質心，就把該客戶分配到那一群去。

3.9 分群結果分析

在分群結束後，本研究還提供一介面讓使用者可以分析群間或群內的交易特性，如圖 2 與圖 3。在此我們利用各群的質心與歐幾里德距離來計算相似的程度，為了讓使用者更容易理解，我們把相似度定義為 1-正規化後的距離比率，如公式(8)：

$$S = 1 - \frac{RD}{MD}, \quad (8)$$

S 表示相似度， RD 表示實際距離， MD 表示最大距離。

例如共有 10 個品號，2 組分群間距離最大的交易組合為 0000000000 與 1111111111，因此距離約為 3.1623；接著我們算出實際上 2 群間的距離並加以正規化，就可以得到此距離佔總距離的比例。例如計算出來為 1.0123，則 2 群間的距離比為 $(1.0123 \times 100) / 3.1623 = 32.01\%$ ，也就是 2 群間的相似程度為 67.99%。

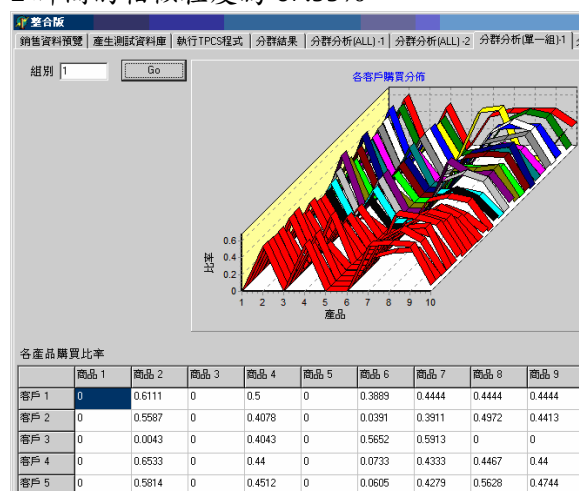


圖 2: 顧客購買產品分析

整合版									
銷售資料預覽 產生測試資料庫 執行TPCS程式 分群結果 分群分析(ALL)-1 分群分析(ALL)-2 分群分析(單一組)-1									
群內各客戶相似程度 (b距離)									
	客戶 1	客戶 2	客戶 3	客戶 4	客戶 5	客戶 6	客戶 7	客戶 8	客戶 9
客戶 1	100%	86.45%	71.26%	88.2%	86.98%	88.13%	83.85%	85.32%	84.47%
客戶 2	86.45%	100%	66.5%	95.97%	96.3%	96.8%	94.13%	96.05%	94.9%
客戶 3	71.26%	66.5%	100%	66.43%	65.16%	67.4%	68.17%	68.44%	62.91%
客戶 4	88.2%	95.97%	66.43%	100%	95.45%	96.38%	92.09%	94.16%	94.8%
客戶 5	86.98%	96.3%	65.16%	95.45%	100%	95.35%	91.1%	93.8%	94.7%
客戶 6	88.13%	96.8%	67.4%	96.38%	95.35%	100%	93.06%	95.32%	94.1%
客戶 7	83.85%	94.13%	68.17%	92.09%	91.1%	93.06%	100%	94.53%	91.3%
客戶 8	85.32%	96.05%	68.44%	94.16%	93.8%	95.32%	94.53%	100%	91.9%
客戶 9	84.47%	94.9%	62.91%	94.8%	94.7%	94.1%	91.32%	91.9%	100%
客戶 10	85.26%	97.18%	66.03%	95.09%	95.67%	95.47%	93.57%	94.97%	95.9%

各客戶與質心距離		
	距離	密度
客戶 1	5.4772	0.3043
客戶 2	15.3881	0.086
客戶 3	14.7374	0.0641
客戶 4	14.3069	0.0954
客戶 5	17.055	0.0793
客戶 6	20.8074	0.0663

圖 3: 顧客相似程度分析

3.10 建立網路模型

因 BPN 具有學習精準度高與回想速度快的優點，也相當適合做分群的研究，因此我們以客戶的交易資料與最後的分群結果為訓練資料，並利用 PCNeuron[4]這一套工具來建構 BPN，以達到快速分群的目的。在此輸入層的處理單元集合為 $T=\{X_1, X_2, \dots, X_m\}$ ， m 為用來分析交易的總產品數。輸出層的處理單元集合為 $O=\{Y_1, Y_2, \dots, Y_k\}$ ， k 為最終的分群數量。

隱藏層以 1-2 層最恰當，因 1-2 層已足夠反應大多數的狀況，太多層只有增加複雜度。經測試我們用一層的隱藏層就有相當不錯的結果，因此隱藏層只設一層。而隱藏層的單元數量會根據每次分析的產品數量與最終的分群結果不同而改變，在此定義為：

隱藏層的單元數 =
 (輸入層的單元數 + 輸出層的單元數) / 2 (9)

起始的權重區域設為 0.3，並以亂數 0.456 隨機產生權重，因此兩個處理單元間的權重 ω_{ij} 的值為正負 0.3 之間隨機均佈的亂數，在此權重區域與亂數均可由使用者自定。訓練次數先設定為 50。學習速率 (Learning Rate) 一般設為 -1 或 0.5 是較好的，經實驗測試學習速率需在 0.5 以上會有較好的效果，設定畫面如圖 4 所示。

處理單元數目	19	產生訓練檔
隱藏層1單元數目	11	產生網路設定檔
輸出單元數目	3	Training and recalling
訓練範例數目	50	看結果
測試範例數目	1	
學習循環數目	50	
加權值區域	0.3	
加權值亂數種子	0.456	
learning rate	0.7	

圖 4: BPN 訓練參數設定

4. 實驗結果與討論

4.1 虛擬資料測試

在實驗的部分我們先以虛擬的 51 筆交易資料來測試，其中顧客數為 17，購買商品數為 27 個，預期分群結果為 5 群，測試結果得到的分群數為 7 群。

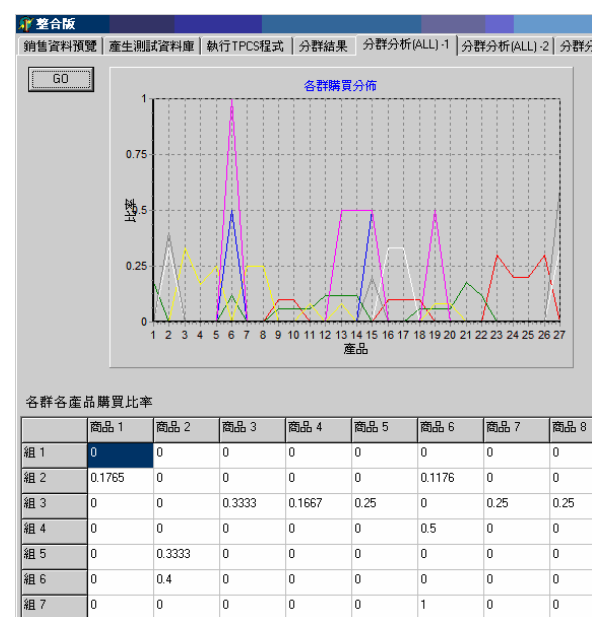


圖 5: 各群購買產品分佈

我們也可以檢驗某個分群裡的客戶，是不是真的適合分配到此分群。我們由圖 6 可以看到，這 5 個客戶與所屬的分群 2 的質心最為接近，也就是最相似，因此我們可以判斷此分群的正確率為 100%。

整合版							
銷售資料預覽 產生測試資料庫 執行TPCS程式 分群結果 分群分析(ALL)-1 分群分析(ALL)-2 分群分							
與其他分群相似程度(by距離) 正確率 100%							
	分群 1	所屬分群(2)	分群 3	分群 4	分群 5	分群 6	分群 7
客戶 1	86.26%	94.34%	85.17%	83.44%	84.08%	82.79%	72.51%
客戶 2	86.87%	94.54%	86.11%	83.64%	84.29%	82.98%	72.41%
客戶 3	83.28%	88.43%	84.7%	84.96%	81.86%	80.71%	79.71%
客戶 4	82.68%	87.57%	82.21%	86.39%	81.3%	80.19%	80.75%
客戶 5	70.75%	76.85%	70.47%	69.57%	69.91%	69.21%	61.51%

圖 6:分群內的客戶與其他分群的相似度

4.2 實際資料測試

在本研究中我們針對三種不同類型的企業做分群分析，希望能驗證 TPCSNN 分群方法的正確性。為維護企業資料隱密性，對於所有實際資料均轉換成代號，這三家企業資料如下：

表 6:企業類型

企業類型	經營項目	客戶數量	品號數量
白米公司	米食之生產與銷售	10564	268
糖果貿易商	日本糖果、餅乾、巧克力等休閒食品代理	1179	1958
家電代理商	代理銷售日本系列家電產品	1990	9305

實際資料驗證結果整理如表 7，我們針對每一群去分析其正確率，看顧客被誤判的機率有多高，表內的組數就是最終的分群結果。

表 7(a):白米公司分群結果

白米公司				
分析品號:25 分析客戶:50 分析日期區間:3 個月 交易筆數:2189				
群間相似程度	分群	組 1	組 2	組 3
	組 1	100.00%	79.04%	58.48%
	組 2	79.04%	100.00%	71.30%
	組 3	58.48%	71.30%	100.00%
群內正確率		89.96%	100.00%	100.00%

表 7(b):糖果貿易商分群結果

糖果貿易商						
分析品號:50 分析客戶:100 日期區間:3 個月 交易筆數:1354						
群間相似程度	分群	組 1	組 2	組 3	組 4	組 5
	組 1	100.00%	79.58%	76.78%	86.94%	78.42%
	組 2	79.58%	100.00%	70.91%	77.52%	67.00%
	組 3	76.78%	70.91%	100.00%	70.21%	66.91%
	組 4	86.94%	77.52%	70.21%	100.00%	71.01%
	組 5	78.42%	67.00%	66.91%	71.01%	100.00%
群內正確率		90.07%	100.00%	100.00%	100.00%	100.00%

表 7(c):家電代理商分群結果

家電代理商				
分析品號:50 分析客戶:100 日期區間:3 個月 交易筆數:4633				
群間相似程度	分群	組 1	組 2	組 3
	組 1	100.00%	93.32%	93.74%
	組 2	93.32%	100.00%	89.79%
	組 3	93.74%	89.79%	100.00%
群內正確率		97.87%	100.00%	100.00%

4.3 產生網路模型

我們利用前面步驟產生的分群結果來訓練 BPN，之後當有一個新的客戶想要做分群時，只需將此客戶有購買所欲分析產品的交易資料輸入，就可以快速得到分群的結果。我們以白米公司的資料做驗證，表 8 為該公司分群後各群購買各產品的機率：

我們模擬第二群客戶的資料分佈，輸入 2 筆測試資料驗證，並預期應該會落在第二群。並以之前的分群結果做為訓練資料，結果如預期般的落在第二群之中，如圖 7 所示，且判斷時間只需數秒鐘。

表 8: 白米公司各群的購買機率分佈

各組各商品購買機率			
組別 商品	組 1	組 2	組 3
1	0.0524	0.3465	0.963
2	0.4178	0	0
3	0.2581	0.099	0
4	0.0078	0.1386	0
5	0	0.2475	0
6	0.2475	0.1683	0
7	0.0262	0.3267	0.7778
8	0.0204	0.1485	0
9	0.0238	0.0693	0
10	0.1087	0.0198	0
11	0.0068	0	0
12	0.0213	0	0.6667
13	0.3658	0.3168	0
14	0.0344	0.3564	0.8148
15	0.0844	0.2277	0.4444
16	0.3372	0.1782	0.2222
17	0.3857	0	0
18	0.3338	0.1287	0
19	0.1339	0.198	0.0741

測試資料分群結果(屬於該分群的機率)

	分群 1	分群 2	分群 3
測試組 1	0.198	0.585	0.137

圖 7: 模擬資料的分群結果

4.4 實驗結果討論

經過實際資料驗證，在重複的實驗中有些值得注意的地方提出來做討論。

4.4.1 TPCSNN 計算的時間成本

本研究所採用的 TPSCNN 計算方法必須重覆運算大量的資料，例如當選擇 n 個品號來分析時，就會有 $2^n - 1$ 種規則需要去計算，而且每當有分群有修正時就必須重複計算一次，因此耗費的時間成本相當可觀。以 P4 3.0 的 PC 測試，如果是 50 個品號大約需要一天的時間

程式才能跑完；如果是 200 個品號，約需五天的時間。這是未來可以再研究及改進的地方。然而，執行時間雖然長，但公司僅需耗費一次的時間成本來分析即可。

4.4.2 顧客交易資料的過濾

因為 TPCSNN 計算時間過長，且一般企業的品號數量都很多，因此本研究希望能以企業最關注的品號來做分析。在此我們選用企業銷售次數最多的產品來分析而不是用單價，因為貴的東西可能客戶數本身就少，無法測試分群的效果。另外本研究的資料來源是選擇過濾 ERP 的銷貨單，像贈品或運費等品號，並不是我們所要分析的，但因為這樣的品號在銷貨單上出現的次數會相當多，為了避免混淆了分群結果，這樣的資料也必須過濾掉。

4.4.3 交易資料特性對分群效果的影響

研究實際分群後的結果，有時會發現同一群內客戶之間的購買特性不一定都很相似，因為在實際資料中，每個客戶購買的規則以及這些規則的交易數量都相當多，加上我們選擇的都是最重要的品號與客戶，因此可能會有一種狀況，就是這家企業有幾家主要的交易客戶，這些客戶佔了大多數的交易。這種客戶的交易資料很容易與其他客戶相似，造成以此客戶為中心的群聚效應，例如 A 與 B 相似，A 又與 C 相似，但 B 與 C 不相似，但因為根據計算，他們都應該與 A 歸為一群，因此 B 與 C 就會被歸在一起，在分析時就會讓人混淆。

另外，當交易規則數量過多時，程式也無法做很正確的分群。例如 A 與 B 各有 5 筆交易資料與 5 種交易規則，但 A 與 B 的交易規則卻只有一筆相同。在這種情況下，因為 TCE 計算時是相關係數較高的 2 群會合併在一起，即使 A 與 B 只有一個關聯規則是相同的，程式算出來就有可能造成 TCE 值較高，因此 A 與 B 還是會合併成一群，但這是合理的結果。

4.4.4 類神經網路分群方法驗證

本研究應用類神經網路分群方法進行分群結果分析與學習，經實驗驗證，訓練出來的類神經網路正確性相當良好，且藉由這樣的方式，我們可以避免分群計算時間過長的問題，如果有一個新客戶需要分析時，只需將其交易資料輸入系統，就可得到其分群結果，這可幫助使用者更能接受 TPCSNN 這個分群方法。

5. 結論

本研究是植基於一個以顧客過去的交易特徵為基礎的分群方法，另外加入顧客交易資料萃取的商業邏輯，以及對分群結果的分析與解釋，進一步應用類神經網路的方法，讓使用者可以不用再花大量的時間成本，就可以得到一個新顧客所屬的分群。經過模擬資料與實際資料的驗證，證明這樣的分群方法在處理含有商業目的分群時，有相當不錯的效果。

參考文獻

- [1] 林育臣，(2002) “群聚技術之研究”，朝陽科技大學資訊管理研究所，碩士論文。
- [2] 林惠玲，陳正倉，(1999)，應用統計學，雙葉書廊，台北市，第 529-535 頁。
- [3] 郭盛陽，(2004) “以顧客交易特徵為基礎的顧客分群方法”，朝陽科技大學資訊管理研究所，碩士論文。
- [4] 葉怡成，(2004)，類神經網路模式應用與實做。
- [5] Ester, M., Kriegel, H. P., Sander, J., and Xu, X. (1996), “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pp. 226-231.
- [6] Kotler, P. (1994), *Marketing Management: Analysis, Planning, Implementation, And Control*, 8th ed., Prentice Hall, Inc.
- [7] Kotler, P., (2006)，邁向品牌與全球化之路-科特勒行銷論壇。
- [8] Kotler, P. and Armstrong, G. (1997), *Marketing An Introduction*, 4th ed., Prentice Hall, Inc.