

預測客戶潛在的消費行為---使用以遺傳演算法為基礎的組合多層分類器

陳素雯

屏東永達技術學院 財務金融系

email:tamshui@ebtnet.net

摘要

近年來，隨著網際網路的蓬勃發展，全世界的商業公司及政府單位都透過電子商務(e-commerce)來強化本身的競爭力。其中，商業公司則希望透過資料採掘技術(data mining)來爭取顧客及預測客戶潛在的消費行為來獲取利益。但是各式各樣不同的商業公司其經營環境也均各不相同，實在很難找到一種最適當且最有效率的資料採掘演算法來讓各個不同公司同時接受使用。最近出現一個新的趨勢，就是結合多層分類器(multiple classifiers)來改善資料分類結果。在本篇論文中，我們藉由提出一個以遺傳演算法為基礎的組合多層分類器來預測客戶的潛在消費行為。這個方法通過國內一家領導級的電子商務公司測試及評估，我們也針對一般分類問題使用辨識手寫文字來驗證這個方法正確性無誤。這兩個評鑑結果，均顯示我們所提出來的的方法確實比單一分類器來得有效率。

關鍵詞: 購買行為分析，多層分類器，電子商務

Abstract

With the increasing popularity of the Internet, EC companies are eager to learn about their customers using data mining technologies. But the diverse situations of such companies make it difficult to know which is the most effective algorithm for the given problems. Recently, a movement towards combining multiple classifiers has emerged to improve classification results. In this paper, we propose a method for the prediction of the EC customer's purchase behavior by combining multiple classifiers based on genetic algorithm. The method was tested and evaluated using Web data from a leading EC company. We also tested the validity of our approach in general classification problems using handwritten numerals. In both cases, our method shows better performance than individual classifiers and other known combining methods we tried.

1. 研究動機與目的

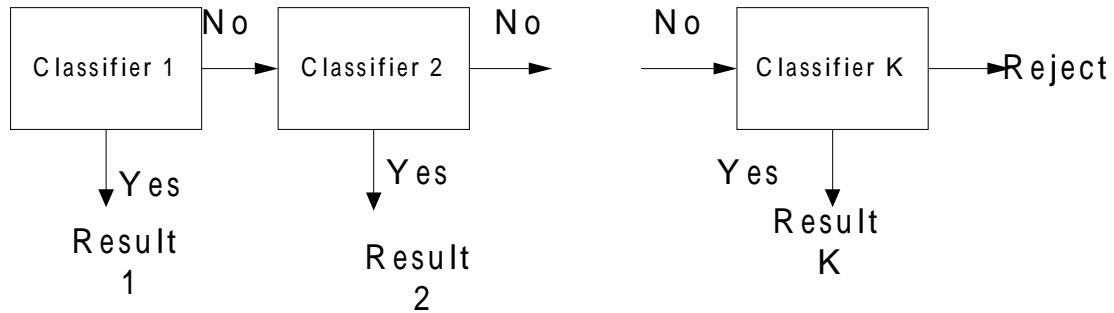
隨著全球資訊網(WWW)的快速發展，電子商務的技術應用讓企業的生產力提昇、作業成本降低、以及作業流程簡化，這使得整個經營變得較以往更有績效，但是電子商務技術的引進真正讓企業經營者心動的，莫過於提供管理和決策時的支援，以增加企業的競爭優勢。對於一些擁有眾多客戶的傳統企業，在進入 e 時代後，在高度競爭的市場中為了增加其產品的市場佔有率或是開發客戶潛在的消費能力，一般

都積極導入先進的「客戶關係管理系統」(customer relationship management, 又簡稱 CRM)為客戶的需求提供快速的解決方案,企圖與每一客戶建立終身的伙伴關係,以確保他們不會輕易轉向競爭對手。然而建立 CRM 必須透過對長期累積的客戶資訊與知識,進行各項分析、統計、並由錯綜複雜的資訊中發覺個別客戶的潛在消費需求,然後再根據特定的行銷訊息提供給客戶。傳統的大眾行銷常以地域和人口統計為分析的重點,然而根據研究指出,這種預測客戶消費行為的有效性不到 5%。但隨著資料倉儲(data warehouse)的興起,很多相關的應用技術也跟著發展起來,其中「資料挖掘」(data mining)的應用就是運用人工智慧(AI),基因演算法(GA),及統計分析等相關技術,企圖從排列方式雜亂無序、未結構化的資料中找出規則並整理成有用的資訊。所以如何應用資料挖掘技術來預測客戶消費行為是一項值得研究的工作。在本研究中,我們提出一個以基因演算法為基礎的多層分類器來預測客戶潛在的消費行為。

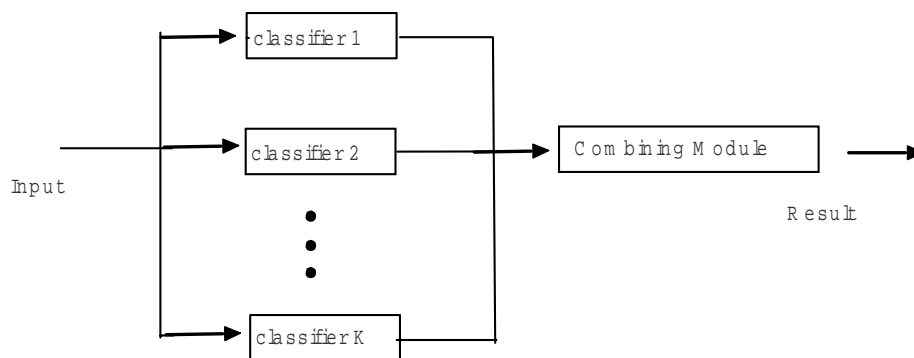
從資料挖掘的觀點來看,預測客戶潛在的消費行為可以被視為一個分類問題,而分類問題本身就是資料挖掘研究領域的一個重要工作。目前已有許多技術及演算法來處理分類問題。有時候,針對某一特定問題時,某一種演算法會比其它種的演算法來的好用,但是不可能存在某一種演算法的效率(performance)可在不同的環境狀況下,皆勝過其它不同的演算法的效率。通常選擇演算法需要根據問題的特徵來決定,可是當遇到問題時,往往無法事先知道問題的徵結,所以也就無法選擇最佳的技術及演算法。

許多研究都已了解到使用單一分類技術(single classification technique)存在著許多限制。這個研究發現最近啟發了很多使用多層分類器來求得更佳效能的研究[1,3,4,5,6,7,9,10,13,14,16,18,19]。使用多層分類器的好處及特色早在國際上的手寫辨識研究領域上被證明承認[10,13,16]。

組合式的多層分類器有兩類:第一類是序列式的組合多層分類器(serial combination),第二類是平行式的組合多層分類器(parallel combination)。序列式的組合多層分類器是將分類器依線性方式一個接一個串接在一起,而它的結果取得是由前一個分類器傳到下一個分類器[3,7]。平行式的組合多層分類器則將分類器平行式地接在一起,當輸入一個值時,多層分類器同步地處理這些輸入值,這些分類器產生的結果透過一個組合演算法整合起來[1,6,9]。在序列式的組合例子中,這個序列組合的順序對整個分類器的績效產生決定性影響,其各別的分類器反而對整個序列組合沒有太大影響。在平行式的組合例子中,整個系統的績效則僅與其組合演算法有關。在本篇研究裏所描述的分類器則是平行式的分類器。序列式和組合式分類器結構圖分別在圖 1 和圖 2 內所示。



圖一：序列式的組合多層分類器圖



圖二：平行式的組合多層分類器圖

一些在組合分類器中常用到的方法包括有: Bayesian, BKS,及 Borda Count。除此之外有一些工作也用到類神經網路及模糊邏輯演算法[1,16,19]。因為每一個單獨分類器的輸出結果都將成為下一個組合分類器的輸入源，也因此分析分類器能支援什麼樣種類的輸出資訊就很重要。各式各樣分類器的輸出資訊能分為三大類: 抽象類 (abstract level)、層級類 (rank level)、評估衡量類(measurement level)[18]。抽象類的分類器僅輸出類別標識(class label)，層級類的分類器則輸出每一個類別的層級，而評估衡量類的分類器則對每個類別指定一個評估值來指示說明和類別有關的輸入樣式機率。類神經網路則是評估衡量類分類器的代表性範例。評估衡量類分類器是比抽象類及層級類分類器更能提供更豐富資訊。在本篇論文中，為了結合像類神經網路的評估衡量類分類器，我們提出一個以基因演算法為主的方法。

底下我們就本篇論文的組成架構作一說明。第二節說明目前文獻中屬於平行式組合分類器中大部份被使用的方法。在第三節中則提出我們的理論架構---以基因演算法為主的多層分類器組合架構。而在網站上的實驗結果及手寫數字驗證則在第四節中介紹。結論和未來研究則放在第五節中。

2. 文獻探討

各式各樣的分類器組合方法及架構早已被提出來了而它們的實用之處也被實驗

證明, Xu[18]等人企圖藉由將貝氏正規化(Bayesian formalism) 投票法則(voting principle) 及丹波斯特-雪佛爾(Dempster-Shafer)理論等方法來將個別的分類器組合起來。 Ho 等人[4]建議使用以 Borda Count 和邏輯式的回歸方法來將個別分類器輸出組合減少及重新排序。在本節剩下的篇幅, 我們介紹五種相當有名的組合方法: 多數決投票法, 貝氏法, BKS, Borda count, 以及類神經網路。

2.1 多數決投票法則(majority voting)

投票法是一種最常用來結合超過一種決擇的方法。目前有各式各樣的投票策略諸如全場一致策略, 多數決策略, 及 Borda Count。這個多數決策略方法所產生的決定乃根據是否全場一致或是至少超過一半分類器同意而定。當沒有超過一半同意時則拒絕這個輸入。這個方法相當簡單且不需額外記憶體。然而這個有一個缺點是所有的分類器均被一視同仁而不在乎各別的分類器是否有特殊不同。

2.2 貝氏法(Bayesian)

投票法則只考慮每個分類器的結果, 而使用貝氏法則(Bayesian formalism)卻考慮每個分類器的錯誤。現在假定有 M 個類別, 其編號分別從 1 到 M , 那麼第 k 個分類器錯誤, 則可以用一個二維陣列矩陣來表示, 其中 $k=1, \dots, K$, 如下所示。

$$PT_k = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1M} \\ n_{21} & n_{22} & \cdots & n_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ n_{M1} & n_{M2} & \cdots & n_{MM} \end{bmatrix} \quad (1)$$

在矩陣中的列(rows)所代表的意義是真實輸入值而行(columns)標識所代表的是每一個分類器的分類結果。在選擇類別 j 的分類器條件下, 其中 $1 \leq j \leq M$, 當類別 i 為真的情況發生時, 其中 $1 \leq i \leq M$, 其機率為:

$$P(x \in C_i | e_k(x) = j) = \frac{n_{ij}(k)}{\sum_{i=1}^M n_{ij}(k)} \quad (2)$$

其中當真實類別對於輸入值為 x 時, $e_k(x)$ 是分類器 k 所選擇的一個類別標識。

對類別 i 而言, 其信賴函數(belief function)可將其表示為所有條件機率的總合, 如下所示:

$$BEL(i) = \eta \prod_{k=1}^K p(x \in C_i | e_k(x) = j), \text{ for } i=1, \dots, M \quad (3)$$

其中 η 是一個滿足 $\sum_{i=1}^M BEL(i) = 1$ 的正規化係數。

信賴函數 $BEL(i)$ 是類別 i 的所有分類器之乘積，而且它展示出類別 i 的全部正當性。最後對決定值就取最大的 BEL 值的類別標識。這個組合規則如下所示：

$$F(x) = \begin{cases} j & \text{if } BEL(j) = \max_{i \in A} (BEL(i)) \wedge BEL(j) \geq \alpha (0 < \alpha \leq 1) \\ reject & \text{otherwise} \end{cases}$$

2.3 行為認知空間法(Behavior-knowledge space method)

貝氏法有一個非常重要的限制那就是它需要多層分類器內的分類器彼此互相獨立，這種情況在現實應用中是很少見成立的。有一種方法可用來克服這種限制就是行為認知空間法(behavior-knowledge space, 簡稱 BKS)[5]。一個 BKS 就代表一個 k -維空間，其中每一個維度都對應到一個分類器的決定值。表一就代表一個 2 維 BKS 的例子。其中表內的行與列分別代表 2 個分類器(分類器 1 和分類器 2)的個別決定值 我們也假設這些決定值從 1 到 11。

因此在表 1 中的每一個儲存格的值都代表著個別分類器彼此交集所產生的決定值也是 BKS 方法的基本計算單元。一般來說，在一個 k 維的 BKS 表格中，每一個儲存格內含的值可以表示成 $BKS(e(1), \dots, e(K))$ ，其中分類器 1 的決定值為 $e(1)$ ，而分類器 K 的決定值為 $e(k)$ ，依此類推。要將每一個分類器的決定值結合起來，BKS 採兩階段作法：學習階段及決定階段。在學習階段中，先對 K 個分類器產生決定階段所需的 k -維 BKS 表，進入決定階段後，再用下列的法則來決定最後結果：

$$F(x) = \begin{cases} R_{e(1)\dots e(K)}, & \text{if } T_{e(1)\dots e(K)} > 0 \wedge \frac{n_{e(1)\dots e(K)}(R_{e(1)\dots e(K)})}{T_{e(1)\dots e(K)}} \geq \lambda \\ reject & \text{otherwise} \end{cases}$$

其中 λ 是一個轉折值($0 \leq \lambda \leq 1$)，它控制著最後決定值的信賴度，在 $BKS(e(1), \dots, e(K))$ 中， $R_{e(1)\dots e(K)}$ 是最好的代表值，而 $T_{e(1)\dots e(K)}$ 是在 $BKS(e(1), \dots, e(K))$ 中全部的輸入樣本數， $n_{e(1)\dots e(K)}$ 是在 $BKS(e(1), \dots, e(K))$ 中類別 m 的全部輸入樣本數。

表 1: 二維的行為認知空間

e(1)/e(2)	1	...	j	...	11
1	(1,1)	...	(1,j)	...	(1,11)
⋮	⋮	⋮	⋮	⋮	⋮
i	⋮	⋮	(i,j)	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮

11	(11,1)	...	(11,j)	...	(11,11)
----	--------	-----	--------	-----	---------

2.4 Borda count

Borda count 是將多數決投票法則(voting majority)一般化，它是一個很有用的群組一致性函數。對某一個類別而言，Borda count 是計算每個分類器其排在該類別之下的所有類別的數目之和。如果現有 k 個分類器，就第 j 個分類器來看，假設 $B_j(i)$ 是排在類別 i 之下的所有類別的數目之和，則對類別 i 而言，borda count 可被定義如下：

$$B(i) = \sum_{j=1}^K B_j(i) \quad (6)$$

一致性排序的取得可藉由將這些類別排好使得 borda counts 的排列方式由大排到小(descending)而且擁有最大的 borda count 的類別標識能被挑選出來。要實作 Borda count 這個方法是很容易的但是在將來發展的潛力上它並沒有考慮到個別分類器的差異性。這裏有一個補救方法，那就是將每一個分類器產生的排序分數加權計分，此種方法又稱為加權 borda count 方法。

2.5 透過類神經網路的組合

這個方法是使用類神經網路的學習能力來將分類器結合起來。Huang[6]提出一個組合方法，它使用資料轉換及「通用差距法則」(general delta rule)。每一個分類器的輸出結果均先被轉換成一種評估可能性的形式。這個被轉換的評估值被輸入到類神經網路的輸入層然後再由類神經網路產生一個最後的分類結果。

3. 以遺傳演算法為基礎的多層分類器

在此節中，我們提出一個以遺傳演算法為基礎的多層分類器方法，此方法將多層分類器所產生的多個分類評估結果整合成單一結果。現考慮一個圖形辨識分類問題，其中圖樣(pattern) x 來自一組 N 個類別 $C_1, C_2, C_3, \dots, C_N$ 之中。令集合 $\Lambda = \{1, 2, 3, \dots, N\}$ 是一組類別標識。假定我們有 K 個分類器，每個分類器均表示一個事件---給定圖樣 x 伴隨著一組評估向量 $M_k = \{m_{1k}, m_{2k}, \dots, m_{Nk}\}$ ，其中 $k = 1, \dots, K$ ，而且 m_{ik} 是類別 i 的第 k 個分類器的評估值。令 $W_k = \{w_{1k}, w_{2k}, \dots, w_{Nk}\}$ 是第 k 個分類器對所有類別的加權向量(weighted vector)，它代表著第 k 個分類器對所有類別的相對重要性。 w_{ik} 代表是類別 i 對第 k 個分類器的重要性而且衡量出對第 k 個分類器和其它分類器相比較時有多重要。

現在對類別 i 而言，要取得其輸出值 o_i 時， k 個分類器所供應的評估值 $m_{1k}, m_{2k}, \dots, m_{ik}$ 是來自於相對應的加權值 $w_{i1}, w_{i2}, \dots, w_{iK}$ 。然後類別 i 的輸出值 o_i 是這些加權評估值的總

合。這個式子可寫成如下所示:

$$o_i = \sum_{k=1}^K w_{ik} m_{ik} \quad (7)$$

這個表式法可以寫成矩陣形式。

這個最後決定值可透過挑選擁有最高輸出值 o_i 的類別標識來取得，如下所示:

$$E(x) = \begin{cases} j & \text{if } o_j = \max_{i \in \Lambda} o_i \wedge o_j \geq \alpha \\ \text{reject} & \text{otherwise} \end{cases} \quad (8)$$

其中 α 是一個臨界值。

在本篇論文中所提出的網路架構如圖 3 所示，這個組合架構它有兩層:一層是輸入層另一層是輸出層。從每一個分類器產生的評估值被放進輸入層。而 o_1, o_2, \dots, o_N 是輸出層的結果。輸出層的節點數和全部的類別數之和相同。我們提出來的方法和單層網路最大的不同處在於我們的方法用遺傳演算法來將連接加權(connection weights)最佳化。從隨機產生加權值開始，這個加權值逐漸反應每個分類器的相對重要性。

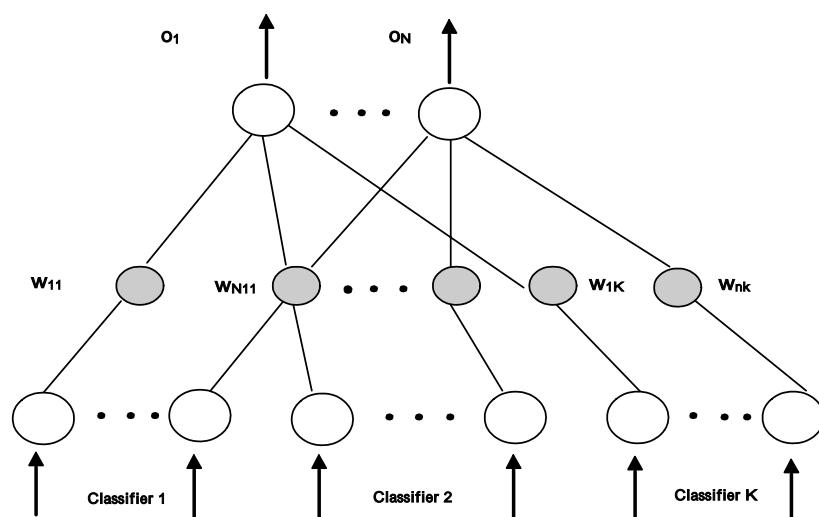


圖 3: 以遺傳演算法為基礎的多層分類器組合模組之結構

3.1 利用遺傳演算法來學習

遺傳演算法是一種利用生物演化過程為基礎的隨機搜尋演算法來求得最佳化的其中一種方法[12]。目前已有許多使用遺傳演算法來求取最佳化的模糊歸屬函數(fuzzy membership function), TSP 等等相關領域的研究[17]。它常常被用在求取某一系統之參數最佳化。遺傳演算法有一個關鍵特色是有時候在極度複雜及不連續空間情況下或是很難用坡度陡降法(gradient descent)來描述的適度情況(fitness landscape)它仍可找到良好的解答。

在遺傳演算法中，問題會被轉成一串 2 進位元的字串，就稱之為基因(chromosome)。在我們的研究中，它被編碼成一串實數值。一開始各別一些可能的候選答案就稱之為初

始樣本(population)。這裏我們要解決的問題是將加權矩陣(weight matrix)最佳化以便能將不同分類器產生的評估值組合起來然後求得最後決定值。我們提出的方法包括一組候選加權矩陣及使用遺傳演算法來自動求得最佳化。在進入演化過程前，我們必需將加權矩陣的字串表示法加以定義，如圖四所示。

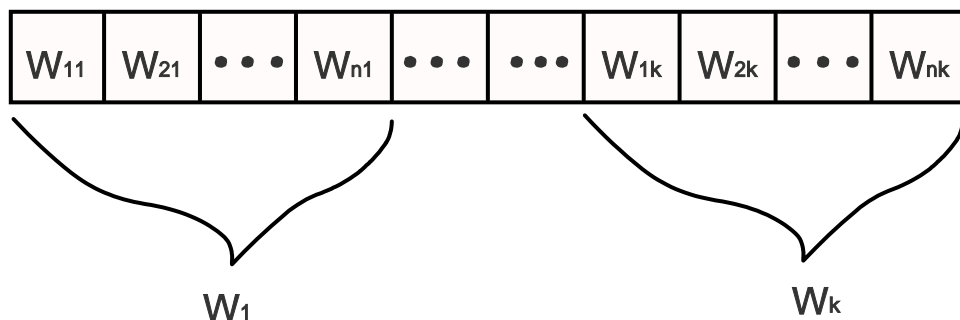


圖 4: 以字串來表示加權矩陣

在圖四中，對一個有 k 個分類器 W_1, W_2, \dots, W_k 的系統裏，加權(weight)可被分成四個部份來表示。 W_k 代表著第 k 個分類器的加權向量。 w_{ik} 則代表著是 W_k 的第 i 個加權值，其意義是對類別 i 而言，對第 k 個分類器的相對重要性。它是一個介於 0.0 和 1.0 之間的正值。遺傳演算法需要一組可實行的初始樣本(population)，然後再透過演化過程修正。這個初始樣本是透過隨機方式設定其加權值。一旦，初始樣本產生後，遺傳演算法就重複地修正這些初始樣本。每重複一次，就表示為產生新一代(generation)，所有新一代的樣本評估來自於最適度函數。新一代的樣本產生是從目前樣本中挑選機率最高之最適當個體來組合。而有些被挑選的個體則維持原封不動到下一代，其它有關新一代的產生方法，則是經由遺傳運算子(genetic operators)來完成，如挑選(selection)，交配(crossover)，突變(mutation)等。

3.2 最適合函數(Fitness function)

任何一種最佳化法則均有一個性能指標(performance index)，當此性能指標最大時即表示到達了最佳解。基因演算法也不例外，我們需事先定義一個適合函數，在基因變化後的每一代族群，都得檢查適合函數的值。譬如每產生新一代群，其適合函數值增大，即表示新一代比上一代優良進步。在本篇論文中每產生一代的過程，當然也需要一個最適合函數(fitness function)，其根據個體的表現好壞來衡量評估打分數。新一代的個體(individuals)就由這些分數來排列順序高低，這些分數就稱為適合值(fitness values)。要對此適合函數做陳述之前，有一個前置工作要做，那就是先定義碰擊函數(hit function)，對每一個候選加權矩陣 WS_q 其碰擊函數定義如下：

$$HF(WS_q) = \begin{cases} 1 & \text{if correctly matched} \\ \left(\frac{o_j(WS_q)}{\sum_{i=1}^N o_i(WS_q)} \right)^\xi & \text{otherwise} \end{cases} \quad (9)$$

其中 $o_i(WS_q) = \sum_{k=1}^K w_{ik} m_{ik}$, w_{ik} 是加權矩陣 WS_q 的一個成員, ξ 是一個常數用來控制學習過程中有關潛在碰擊的影響度, 而 j 則是指輸入時的真實類別。

如果每一個獨立個体能正確地辨別出每一個輸入, 則該分類器的分數就加一分。否則就按真實類別的評估值和所有評估值總合之比的比率來加分, 如式子(10)所示。式子(10)是考慮獨立個體在下一個階段, 其潛在可能碰擊機率。因為它的目的乃在於增加碰擊的比率, 該適合函數就可定義如下:

$$Fitness(WS_q) = \frac{\sum_{i=1}^S HF(WS_q)}{\text{total number of training data}}$$

其中 S 是全部的訓練資料數之總合

3.3 選擇運算子(selection operator)

下一代的樣本(population)的組成是由目前這一代獨立個體根據適度函數來挑選, 我們採用機率模式來挑選獨立個體。挑選候選解答 WS_q 的機率可由式子(11)來求得。

$$P(WS_q) = \frac{Fitness(WS_q)}{\sum_{i=1}^M Fitness(WS_i)} \quad (11)$$

其中 M 代表著樣本的大小。

3.4 交配及突變(crossover and mutation)運算子

在前面的選擇過程中, 被選定的獨立個體僅僅被複製而沒有被改變。為了要產生新一代, 我們會運用交配及突變運算子來處理目前樣本的獨立個體。交配牽涉到將兩個獨立個體結合來產生新的一代。我們採用兩點式的交配, 如圖五所示。切割點是隨機決定。突變運算子隨機地從獨立個體中挑選部份元素出來並稍做修改。這個突變運算確保在長時間運算後加權矩陣的多樣變化性以預防最佳化收斂不動。

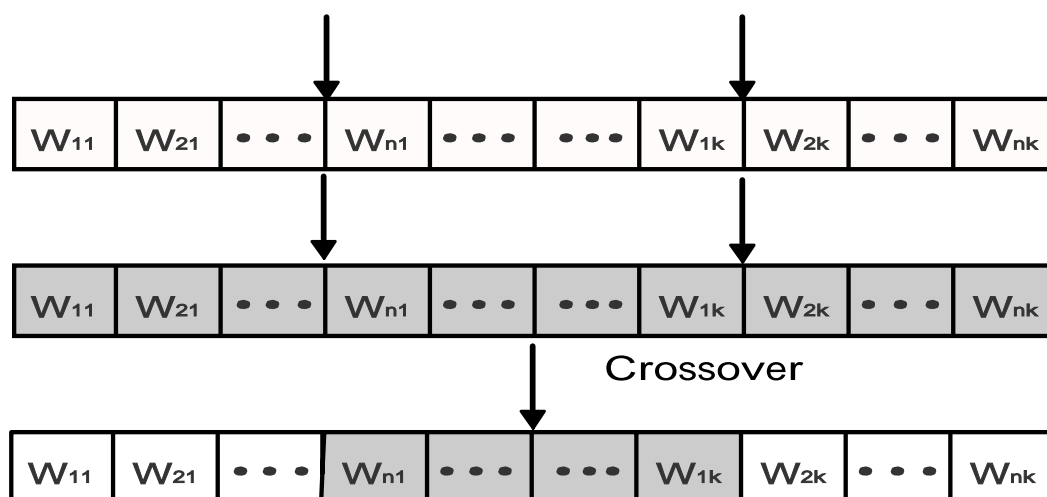


圖 5.兩點式的交配運算子

4. 實驗結果

我們的實驗中有兩組資料。第一，網站資料來自台灣的一家領導級的電子商務公司(奇摩網站:www.kimo.com.tw)，這組資料被用來建立顧客的消費行為預測的模型。我們從資料庫中擷取 15 個特徵包含著 10 個有關人口統計項目如(年齡，性別，教育，職業，婚姻狀態，住址，嗜好，相關單位，客戶類別，及區段號碼)及包含過去一年中，5 個交易特徵如(購買模式，網站使用模式，商品購買頻率，常被購買的項目，常被瀏覽的商品)，總共有 1602 個案例。現有三個類神經網路且擁有不同數隱藏單位 10,20,30(稱為 NN1,NN2,NN3)被用來建立三個不同的分類器並且使用倒遞網路(back-propagation network)來訓練預測對目標商品的潛在購買能力[15]。使用我們所提出來的以遺傳演算法為主的組合方法，從三個分類器所得的結果然後整合成一個統合的預測結果。表 2 記載代表著每一個分類器及我們提出的方法之辨識率。它呈現一個事實，就是我們提出的方法比其它獨立個別的分類器還好。

而有關奇摩電子商務公司所提供資料之實際辨識結果之驗證牽涉到客戶的隱私及公司的商業機密，所以在驗證上有其困難，此部份將留待未來研究進一步克服。

表 2. 每一個分類器及我們提出的方法之辨識率

分類器	分類辨識率	誤判率
K-NN	95.85	4.15
C-NN	94.80	5.20
N-Nn	92.50	7.50
Proposed method	97.80	2.30

第二、為了要評估測試我們提出來的方法的整体表現效率(performance)，我們採用 CENPARMI 這個資料庫。這個資料庫包含了 6000 個未經限制手寫方式的數字，這些數字是美國政府將全美各地方郵局中未經招領的郵件信封上的手寫數匯集而成一個資料庫[20]。這些數字是以 bi-level 格式儲存，目前這個資料庫已廣泛地被全世界學術界及工業界所採用，我們將這 6000 個手寫數字中的 4000 個數字用來訓練獨立個別的分類器。剩下的 2000 個數字中，1000 個用來訓練組合模組，另 1000 個測試組合模組。在圖 6 中，我們列舉了一個 CENPARMI 資料庫中 100 個手寫數字以供參考，在這個資料庫中有一些手寫數字即使是用肉眼也很難辨別的，這也是為什麼 CENPARMI 之誤判率比 KIMO 高很多的原因。

我們選用三個不同類型的類神經網路:K-NN,C-NN,及 N-NN。每一個分類器都是一個多層微感機(multi-layer perceptron)，此多層微感機都是採用倒遞式方式(back-propagation)來訓練。K-NN 使用一個克爾斯遮罩(Kirsh Mask)來偵測邊界[8]。C-NN 使用 chain code 來擷取數字的輪廓[2]。而 N-NN 使用一個 16×16 大小並未經加工的手寫數字原稿。表 3 呈現出這三種分類器和我們提出來的方法的績效比較。在這個實驗中，樣本大小設定為 150 個而突變率為 0.05。獨立個體設定為 100 代。在 40 代後，適合值(fitness value)並無重大改變，而加權向量的收斂速率如圖 7 所示。表 4 比較我們提出的方法和其它的組合演算法的正確率。

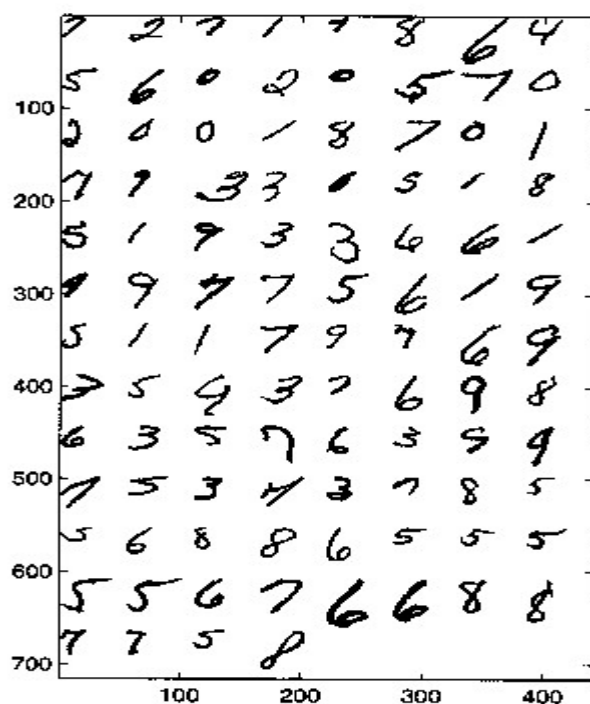


圖 6、在 CENPARMI 資料庫中 100 個手寫數字樣本

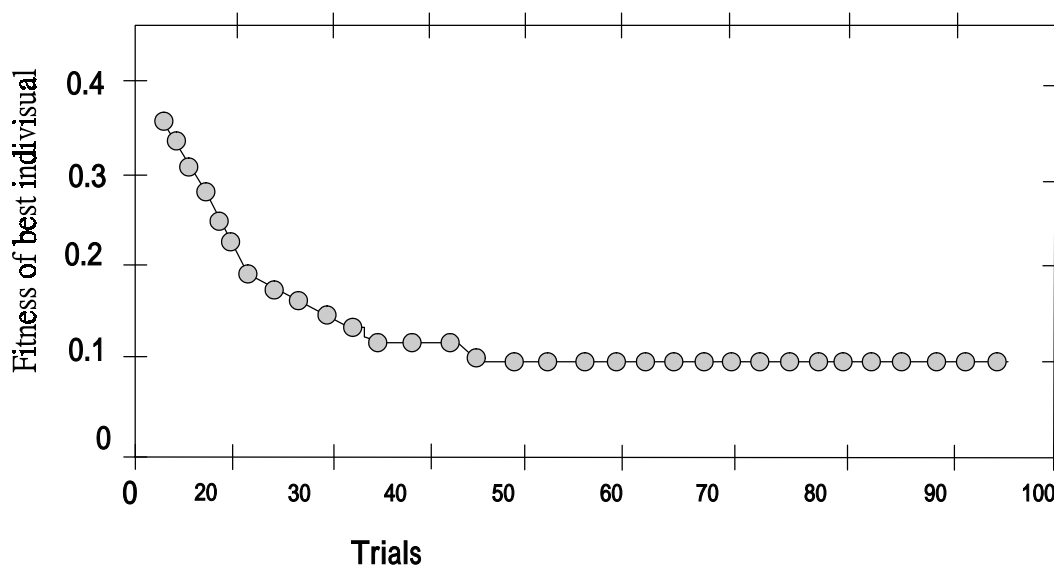


圖 7、加權向量收斂速率

表 3. 每一個分類器及我們提出的方法之分類辨識率

分類器	分類辨識率	誤判率
K-NN	73.1	16.8
C-NN	73.4	16.5
N-Nn	74.1	16.0
Proposed method	76.5	13.5

表 4. 比較我們提出的方法和其它的組合方法

組合演算法	分類辨識率	誤判率
Majority vote	96.65	2.00
Bayesian	97.55	2.45
BKS+Bayesian	97.50	2.50
Borda count	97.40	2.60
Weighted Borda count	97.40	2.60
Condorect	97.40	2.60
Sum of measurements	97.30	2.70
NN	97.60	2.40
Proposed method	97.80	2.30

5. 討論及未來研究

在本論文研究中，我們提出一個以遺傳演算法為主的多層分類器結合方法來預測客戶的消費行為。在我們所提的方法裏，最主要的想法是不同的分類器均有潛能對要被分類的圖樣(patterns)提供相關互補資料。此方法最主要的優點是它有能力結合多層分類器

產生各別決定。我們的實驗資料來自台灣一家電子商務網站(奇摩網站)，實驗結果顯示我們的方法其績效比各別單獨的分類器來得好。這方面也能透過手寫數字的辨識來證明。在第一個實驗中，結果顯示我們所提出的方法能改善預測潛在購買行為的正確性。第二個實驗裏，結果顯示我們所提出的方法比其它組合方法來的好。從這兩個令人鼓舞的實驗結果，可分兩方面總結:第一，這個以遺傳演算法為基礎的組合方法能成功地被用來預測消費者的購買潛力，其預測能力比許多傳統的資料採掘方法還精確。第二，這個方法同時對一般分類問題也是一個有效策略。在我們的未來研究裏，我將運用此篇論文所提的方法到另外的分類問題而且我們也考慮延伸我們的方法來整合不同層次的分類器諸如排序層(rank level)和抽象層(abstract level)等。

參考著作:

- [1] S.B. Cho, J.H. Kim, Multiple network fusion using fuzzy logic, IEEE Trans. Neural Network, 6(2) (1995) 179-501.
- [2] H. Freeman, Boundary encoding and processing, in: B.S. Lipkin, A.Rosenfeld (Eds.), Picture Processing and Psychopictorics, Academic Press, 1970, pp. 241-266.
- [3] P.D. Gader, D. Hepp, B.Forester, TPeurach, B.T. Mitchell, Pipelined systems for recognition of handwritten digits in USPS ZIP codes, Proc. U.S. Postal Service Adv. Technol. Conf., (1990) 539-548.
- [4] T.K. Ho, J.J. Hull, S.N. Srihari, Decision Combination in multiple classifier systems, IEEE Trans. Pattern Anal. March. Intell. 16(1) (1994) 66-75.
- [5] Y.S. Huang, C.Y. Suen, The behavior -knowledge space method for combination of multiple classifiers, Proc. IEEE Conf. CVPR, (1993) 347-352.
- [6] Y.S. Huang, K. Liu, C.Y. Suen, A neural network approach for multi-classifier recognition systems, Proc. of 4th IWFHR, (1994) 235-244.
- [7] F. Kimura, M.Shridhar, Handwritten Numeral Recognition Based on Multiple Algorithms, Pattern Recogn. 24 (10) (1991) 969-983.
- [8] R. Kirsch, Computer determination of the constituent structure of biomedical images, Comput. Biomed. Res. 4(3) (1971) 315-328.
- [9] J. Kittler, M. Hatef, R.P.W. Duin, Combining Classifiers, Proc. IEEE Conf. ICPR, (1996) 897-901.
- [10] T. Matsui, T. Noumi, I. Yamashita, T. Wakahara, M. Yoshimuro, State of the art of handwritten numeral recognition in Japan--the result of the first IPTP character recognition competition , Proc. of the Second ICDAR, (1993) 391-396.
- [11] M. Menta, Data Mining Your Web Site, Digital Press, Butter-worth-Heinemann, Linacre House, Jordan Hill, Oxford OX28DP, UK, 1999.
- [12] T. Mitchell, Machine Learning, The McGraw-Hill, 1221 Avenue of the Americas, New York, NY 10020, USA, 1997.
- [13] T. Noumi, et al., Result of second IPTP character recognition competition and studies on multi-export handwritten numeral recognition, Proc. of 4h IWFHR, (1994) 338-346.
- [14] J.Paik, S. Jung, Y. Lee, Multiple combined recognition system for automatic processing of credit card slip applications, Proc. of the Second ICDAR, (1993) 520-523.

[15] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, in:D.E. Rumelhart, J.L. McClelland (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of cognition, vol. 1, The MIT press. Five Cambridge Center, Cambridge, MA 02142-1493, USA,1986.

[16] H. Takahashi, T.D. Griffin, Recognition enhancementy linear tournament verification, Proc. of the Second ICDAR, (1993) 585-588.

[17] C.H. Wang, T.P. Hong, S.S. Tseng, Integrating fuzzy knowledge by genetic algorithms, IEEE Trans. Evol Comput. 2(4) (1998) 138-149.

[18] L. Xu. A. Krzyzak, C.Y. Suen, Method of combining multiple classifiers and their application to handwritten numeral recognition, IEEE Trans. Syst., Man Cybern. 22(3) (1992) 418-435.

[19] F. Yamaoka, Y. Ku, A. Shaout, M. Shridhar, Fuzzy integration of classification results in handwritten digit recognition system, Proc. of 4th IWFHR, (1994) 255-264.

[20] G.Y.Chen, T.D.Bui, A.Krzyzak, Contour-based handwritten numeral recognition using multivavelets and neural networks, Pattern Recognition vol. 36 ,2003, pp.1597—1604.