

具自適應核形狀參數的徑向基底函數網路

葉怡成

中華大學資訊管理系

陳重志

中華大學資訊管理系

黃冠傑

成功大學資訊工程系

摘要

徑向基底函數網路(RBFN)常用於分類問題，它的核有球形心與半徑二種參數，這二種參數可用監督式或無監督式學習來決定。但它有一個缺點是視所有自變數有同等地位，故分類邊界是圓形，但事實上每一個自變數對分類的影響力不同，分類邊界是應該是橢圓形較合理。為克服此一缺點，本文提出具自適應核形狀參數的徑向基底函數網路，並以監督式學習推導出其學習規則。為證明此一架構優於傳統的徑向基底函數網路，本研究以五個人為的與七個真實的分類例題進行比較。結果顯示，此一架構確實比倒傳遞網路及傳統的徑向基底函數網路更為準確，狀參數值的大小確實能表現出自變數對分類的影響力高低。

關鍵字：半徑基神經網路、監督式學習、核函數、分類



Radial Basis Function Networks with Adaptive Kernel Shape Parameter

I-Cheng Yeh

Department of Information Management, Chung Hua University

Chung-Chih Chen

Department of Information Management, Chung Hua University

Kuan-Chieh Huang

Department of Computer Science and Information Engineering,
National Cheng-Kung University

Abstract

Radial Basis Function Network (RBFN) is usually employed for classification problems, whose kernel has centroid and radius parameters determined with supervised or unsupervised learning. However, it has a shortcoming that it regards each independent variable as the same position; hence, the boundary of classification is circle. But in fact, each independent variable has different influence to the classification, it is more reasonable that the boundary of classification is ellipse. To overcome the shortcoming, we proposed the RBFN with adaptive kernel shape parameters and deduced its learning rule, using supervised learning. To verify whether the architecture is more accurate than conventional RBFN, experiments with five human-made problems and seven real-world problems were conducted. The results showed that this architecture is really more accurate than Back-Propagation Network and conventional RBFN, and the shape parameters can represent the influence of independent variable to classification.

Key words: Radial basis function network, supervised learning, kernel function, classification



壹、前言

傳統上常用統計學的方法做為建構分類或迴歸模型的依據，例如判別分析、迴歸分析，這些方法在許多問題上都具有不錯的效果。然而在面對許多複雜的問題時，這些方法仍有不足，最主要的原因是傳統統計學方法對於非線性系統以及變數間有交互作用的系統較難適用(Yeh 1999; 葉怡成 2006)。近年來神經網路已被視為非常有效的非線性模型建構工具(陳安斌 & 張志良 2001; 陳安斌 & 許育嘉 2004; 駱至中 & 林錦昌 2005; Park et al. 2004)。神經網路是由許多人工神經元 (artificial neuron) 所組成，人工神經元又稱處理單元 (processing element)。每一個處理單元的輸出，成為許多處理單元的輸入。一個神經網路是由許多個人工神經元與其連結所組成，並且可以組成各種網路模式。其中以倒傳遞網路 (Back-Propagation Network; BPN) 與徑向基底函數網路 (Radial Basis Function Network; RBFN) (圖 1) 應用最普遍。

RBFN是Moody與Darken在1989年所提出的(Moody & Darken 1989)。它採用Gaussian函數(圖 2)做為隱藏單元的轉換函數：

$$H_k = \exp(-net_k) \quad (1)$$

$$net_k = \frac{\sum_i (X_i - C_{ik})^2}{\sigma_k^2} \quad (2)$$

其中 H_k 為隱藏層的第 k 個Gaussian單元的輸出值； net_k 為隱藏層的第 k 個Gaussian單元的淨值； C_{ik} 為第 k 個Gaussian單元的核的第 i 個輸入變數的中心值； σ_k 為第 k 個Gaussian單元的核的半徑。

並以線性迴歸方式連接隱藏單元以及輸出單元。RBFN的困難在於決定隱藏單元的數目以及Gaussian函數的形心以及半徑參數(Carse et al. 1995)。其中形心與半徑這二種參數可用監督式或無監督式學習來決定。RBFN的監督式學習與BPN相似，可用最小化誤差平方和與最陡坡降法導出學習規則；而無監督式學習是採用k-means演算法來找出同類樣本的聚類中心做為每個隱藏單元的高斯函數的形心，然後利用最近鄰居法則決定其半徑的適當大小。

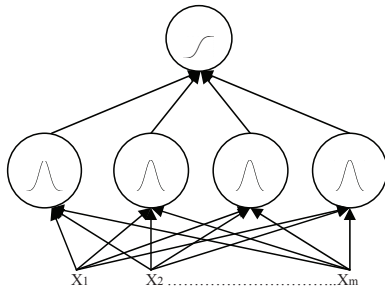


圖1：徑向基底函數網路(RBFN)

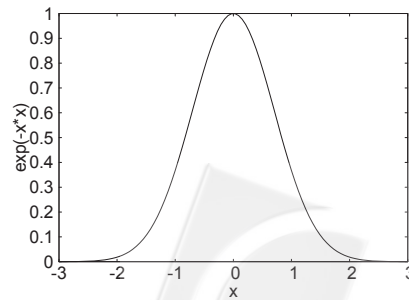


圖2：Gaussian函數網路

RBFN在模式鑑別、時間序列預測以及故障診斷等領域都有廣泛的應用(Whitehead & Choate 1996)，其演算法也有許多不同的變形(Xu 1993; Berthold 1994; Carse et al. 1995; Whitehead & Choate 1996; Cheng 1997; Bugmann 1998; Webb & Shannon 1998; Shibata & Ito 1999; Gomm & Dimg 2000; Han & Xi 2002; Gao & Yang 2002; Park et al. 2004)。但它們有一個缺點是視所有自變數有同等地位，故分類邊界是圓形，但事實上每一個自變數對分類的影響力不同，分類邊界是應該是橢圓形較合理。為克服此一缺點，本文提出具自適應核形狀參數的徑向基底函數網路(Adaptive Radial Basis Function Network; ARBFN)，並以監督式學習推導出其學習規則。為證明此一架構優於傳統的徑向基底函數網路，本研究以五個人為的與七個真實的分類例題進行比較。本文第二節將推導ARBFN的學習規則，第三節與第四節則分別以人為與實際的分類例題進行比較，第五節為總結。

本研究與其它文獻不同之處與創新的貢獻簡述如下：

1. 超橢圓的徑向基底函數

傳統的RBFN視所有自變數有同等地位，故在隱藏層的分類邊界是圓形，再以線性迴歸方式連接隱藏單元以及輸出單元，組成所需的分類邊界。本研究則在隱藏層的Gaussian核函數中加入「變數加權值」，可以調整核函數為任意超橢圓，以匹配不同形狀的分類邊界。此「變數加權值」可視為核函數的形狀參數。本研究也以七個已發表的文獻的例題證明加入形狀參數的核函數確實能大幅提高分類模型的準確度，大幅超越傳統的RBFN，甚至比倒傳遞網路更加準確。

2. 統一的數學理論架構

傳統的RBFN經常以無監督式學習規則來決定核函數的形心、半徑參數，以監督式學習規則來調整隱藏單元到輸出單元之間的連結加權值、輸出單元之門限值，這種混合式、二階段的方法可能無法建構最佳的分類模型。本研究雖然在可調參數上比傳統的RBFN還多了形狀參數，但採用誤差平方和最小化原理統一導出了所有參數的監督式學習規則，包括核函數的形心、半徑、形狀參數、以及隱藏單元到輸出單元之間的連結加權值、輸出單元之門限值。本研究也透過特殊設計的數值例題證明此監督式學習規則所調整的形心、半徑、形狀參數確實能匹配預設的值。

3. 可估計輸入變數重要性的重要性因子

傳統的RBFN無法估計輸入變數對分類的影響力，因此無法了解模型的重要變數為何，使得分類模型缺少解釋能力。本研究獨創的核函數形狀參數可依監督式學習推導出的學習規則調整大小，而形狀參數大小反應了輸入變數對分類的影響力高低。本研究更進一步以數學原理推導出「重要性因子」，並透過特殊設計的數值例題證明「重要性因子」確實可以精確地估計輸入變數的重要程度，使得分類模型具有解釋能力。



貳、理論推導

一、前向傳播 (回想階段)

(一) 計算隱藏層輸出向量 $\{H_1, H_2, H_3, \dots\}$

$$H_k = f(net_k) = \exp(-net_k) \quad (3)$$

$$net_k = \frac{\sum_i V_{ik}^2 (X_i - C_{ik})^2}{\sigma_k^2} = Q_k^2 \sum_i V_{ik}^2 (X_i - C_{ik})^2 \quad k = 1, 2, \dots, N \quad (4)$$

其中 H_k 為隱藏層的第 k 個 Gaussian 單元的輸出值； net_k 為隱藏層的第 k 個 Gaussian 單元的淨值； C_{ik} 為第 k 個 Gaussian 單元的核的第 i 個輸入變數的中心值； V_{ik} 為第 k 個 Gaussian 單元的核的第 i 個輸入變數的加權值，代表輸入變數的重要程度； σ_k 為第 k 個 Gaussian 單元的核的半徑； Q_k 為第 k 個 Gaussian 單元的核的半徑的倒數，採用倒數的原因是為了避免直接使用半徑的平方時，可能出現分母為 0 的問題。

(二) 計算推論輸出向量 $\{Y_1, Y_2, Y_3, \dots\}$

$$Y_j = f(net_j) = \frac{1}{1 + \exp(-net_j)} \quad (5)$$

$$net_j = \sum_k W_{kj} H_k - \theta_j \quad (6)$$

其中 H_k 為隱藏層的第 k 個隱藏單元的輸出值； net_j 為輸出層的第 j 個單元的淨值； Y_j 為輸出層第 j 個輸出單元的輸出值； W_{kj} 為隱藏層的第 k 單元與輸出層第 j 個單元間的連結強度； θ_j 為輸出層第 j 個單元的門限值。

二、後向傳播 (學習階段)

比較網路所推得的「推論輸出值」與訓練範例原有的「目標輸出值」可得網路誤差，類神經網路依照此誤差作為修正連結中的加權值的依據。因為監督式學習旨在降低網路輸出單元目標輸出值與推論輸出值之差距，所以一般以下列誤差函數(或稱能量函數)表示學習的品質：

$$E = \frac{1}{2} \sum (T_j - Y_j)^2 \quad (7)$$

其中 T_j 為訓練範例輸出層第 j 個輸出單元的目標輸出值； Y_j 為訓練範例輸出層第 j 個輸出單元的推論輸出值。

神經網路學習的目的為修正網路上的加權值，使網路誤差函數達到最小值。當每輸入一個訓練範例，網路即小幅調整連結加權值的大小。調整的幅度和誤差函數對該加權值的敏感程度成正比，即與誤差函數對加權值的偏微分值大小成正比：

$$\Delta W = -\eta \frac{\partial E}{\partial W} \quad (8)$$

其中 η 稱為學習速率(learning rate)，控制每次加權值修改的幅度。

(1) 隱藏層與輸出層之間的處理

網路輸出層與隱藏層間之連結加權值之修正量可用偏微分的連鎖率(chain rule)得到：

$$\Delta W_{kj} = -\eta \frac{\partial E}{\partial W_{kj}} = -\eta \cdot \frac{\partial E}{\partial net_j} \cdot \frac{\partial net_j}{\partial W_{kj}} \quad (9)$$

其中

$$\frac{\partial net_j}{\partial W_{kj}} = \frac{\partial}{\partial W_{kj}} (\sum_k W_{kj} H_k - \theta_j) = H_k \quad (10)$$

$$\frac{\partial E}{\partial net_j} = \frac{\partial E}{\partial Y_j} \cdot \frac{\partial Y_j}{\partial net_j} \quad (11)$$

因

$$\frac{\partial E}{\partial Y_j} = \frac{\partial}{\partial Y_j} \left(\frac{1}{2} \sum (T_j - Y_j)^2 \right) = -(T_j - Y_j) \quad (12)$$

$$\frac{\partial Y_j}{\partial net_j} = \frac{\partial}{\partial net_j} (f(net_j)) = f'(net_j) = f'_j \quad (13)$$

將(12)(13)代入(11)得

$$\frac{\partial E}{\partial net_j} = -(T_j - Y_j) \cdot f'_j \quad (14)$$

定義

$$\delta_j \equiv (T_j - Y_j) \cdot f'_j \quad (15)$$

則

$$\frac{\partial E}{\partial net_j} = -\delta_j \quad (16)$$

將(16)(10)代入(9)得

$$\Delta W_{kj} = -\eta \cdot (-\delta_j) \cdot H_k = \eta \cdot \delta_j \cdot H_k \quad (17)$$



同理，輸出單元的門限值修正量：

$$\Delta\theta_j = -\eta \frac{\partial E}{\partial \theta_j} = -\eta \delta_j \quad (18)$$

(2) 輸入層與隱藏層之間的處理

$$\Delta C_{ik} = -\eta \frac{\partial E}{\partial C_{ik}} = -\eta \cdot \frac{\partial E}{\partial net_k} \cdot \frac{\partial net_k}{\partial C_{ik}} \quad (19)$$

將

$$\frac{\partial E}{\partial net_k} = -\delta_k \quad (20)$$

$$\frac{\partial net_k}{\partial C_{ik}} = Q_k^2 V_i^2 (-2(X_i - C_{ik})) \quad (21)$$

將(20)(21)代入(19)得

$$\Delta C_{ik} = -\eta \cdot (-\delta_k) \cdot Q_k^2 V_i^2 (-2(X_i - C_{ik})) = -2\eta \delta_k Q_k^2 V_i^2 (X_i - C_{ik}) \quad (22)$$

同理， ΔQ_k 計算方式如下：

$$\Delta Q_k = -\eta \frac{\partial E}{\partial Q_k} = -\eta \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial Q_k} \quad (23)$$

將

$$\frac{\partial E}{\partial net_k} = -\delta_k \quad (24)$$

$$\frac{\partial net_k}{\partial Q_k} = 2Q_k \sum_i V_i^2 (X_i - C_{ik})^2 \quad (25)$$

將(24)(25)代入(23)得

$$\Delta Q_k = -\eta \cdot (-\delta_k) \cdot 2Q_k \sum_i V_i^2 (X_i - C_{ik})^2 = 2\eta \delta_k Q_k \sum_i V_i^2 (X_i - C_{ik})^2 \quad (26)$$

同理， ΔV_{ik} 計算方式如下：

$$\Delta V_{ik} = -\eta \frac{\partial E}{\partial V_{ik}} = -\eta \sum_k \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial V_{ik}} \quad (27)$$

將

$$\frac{\partial E}{\partial net_k} = -\delta_k \quad (28)$$

$$\frac{\partial net_k}{\partial V_{ik}} = Q_k^2 2V_{ik} (X_i - C_{ik})^2 \quad (29)$$



將(28)(29)代入(27)得

$$\Delta V_{ik} = -\eta(-\delta_k)(Q_k^2 2V_{ik}(X_i - C_{ik})^2) = 2\eta\delta_k Q_k^2 V_{ik}(X_i - C_{ik})^2 \quad (30)$$

三、變數重要性估計

在ARBFN中，令第 i 個輸入變數對第 j 個輸出變數的一次微分如下：

$$\frac{\partial Y_j}{\partial X_i} = \sum_k \frac{\partial Y_j}{\partial net_j} \frac{\partial net_j}{\partial H_k} \frac{\partial H_k}{\partial net_k} \frac{\partial net_k}{\partial X_i} = \sum_k f'_j W_{kj} f'_k Q_k^2 V_{ik}^2 2(X_i - C_{ik}) \quad (31)$$

其中

$f'_j, f'_k, 2(X_i - C_{ik})$ 為與輸入變數有關的因子，在此略去，只保留與輸入變數無關的因子得到

$$\frac{\partial Y_j}{\partial X_i} \propto \sum_k W_{kj} Q_k^2 V_{ik}^2 \quad (32)$$

令第 i 個輸入變數對第 j 個輸出變數的重要性因子定義如下：

$$S_{ij} \equiv abs \sum_k W_{kj} Q_k^2 V_{ik}^2 \quad (33)$$

此一因子可以用來評估第 i 個輸入變數對第 j 個輸出變數的重要性大小。

參、數值例題

本節以五個人為設計的題目進行實驗，以比較ARBFN相對於RBFN與BPN的效能。為了檢驗學習的效果，取五分之一的資料做測試範例，其餘做訓練範例，以測試範例的預測結果來評估各方法的準確性。

在使用神經網路建模時，最重要的參數為學習速率、網路結構與學習循環。網路架構是指隱藏單元的數目(一般為二個到數十個)，網路結構太小可能無法建立足夠複雜的模型來配適資料，但太大也可能產生過度配適現象。學習循環是指將資料集反覆學習的次數，其值太小可能無法建立足夠複雜的模型來配適資料，但太大也可能產生過度配適現象。本研究的網路的參數設定原則如下：(1) 學習速率：採用在每一個訓練循環完畢即將學習速率乘以一個小於1.0的係數的方式，逐漸縮小學習速率，但不小於預設的學習速率下限值。本研究學習速率初始值=0.1，折減係數=0.95，下限值=0.01。(2) 隱藏單元數目：考慮4, 8, 16個，並以嘗試錯誤方式決定每個網路的最佳隱藏層單元數目。(3) 學習循環設為30000次，然後取誤差均方根最小的學習循環做為採用的學習循環。

例題1. 二維四球分類問題

為探討前節的Gaussian核的形心、半徑、形狀參數學習規則是否真能正確找出應有的值，本研究設計了一個「二維四球分類問題」，其分類公式如下：

$$\begin{aligned}
&\text{當 } (x+0.5)^2 + (y+0.5)^2 < 0.4 \quad \text{或} \quad (x+0.5)^2 + (y-0.5)^2 < 0.4 \quad \text{或} \\
&\quad (x-0.5)^2 + (y+0.5)^2 < 0.4 \quad \text{或} \quad (x-0.5)^2 + (y-0.5)^2 < 0.4 \\
&\text{則屬第一類，否則為第二類。}
\end{aligned} \tag{34}$$

在值域 $-1 < x < 1$, $-1 < y < 1$ 的範圍內，以隨機取點方式取得500筆練習資料。此問題的範例分佈狀況如圖3所示，其中屬第一類的範例形成各自以 $(-0.5, -0.5)$, $(-0.5, 0.5)$, $(0.5, -0.5)$, $(0.5, 0.5)$ 為中心點，半徑為0.4的四個完全封閉的圓形。屬於第二類的範例則是分佈於四個圓形以外的地方。

例題2. 二維四球分類問題：不同形心

為探討前節的Gaussian核的形心參數學習規則是否真能正確找出應有的值，在此設計了一個與前節相似的題目，但其形心位置不同，其分類公式如下：

$$\begin{aligned}
&\text{當 } (x-0.0)^2 + (y+0.5)^2 < 0.34 \quad \text{或} \quad (x+0.5)^2 + (y-0.0)^2 < 0.34 \quad \text{或} \\
&\quad (x-0.5)^2 + (y-0.0)^2 < 0.34 \quad \text{或} \quad (x-0.0)^2 + (y-0.5)^2 < 0.34 \\
&\text{則屬第一類，否則為第二類。}
\end{aligned} \tag{35}$$

此問題的範例分佈狀況如圖4所示。

例題3. 二維四球分類問題：不同半徑

為探討前節的Gaussian核的半徑參數學習規則是否真能正確找出應有的值，在此設計了一個與前節相似的題目，但其半徑大小不同，其分類公式如下：

$$\begin{aligned}
&\text{當 } (x+0.5)^2 + (y+0.5)^2 < 0.6 \quad \text{或} \quad (x+0.5)^2 + (y-0.5)^2 < 0.3 \quad \text{或} \\
&\quad (x-0.5)^2 + (y+0.5)^2 < 0.2 \quad \text{或} \quad (x-0.5)^2 + (y-0.5)^2 < 0.4 \\
&\text{則屬第一類，否則為第二類。}
\end{aligned} \tag{36}$$

此問題的範例分佈狀況如圖5所示。

例題4. 二維四球分類問題：不同形狀

為探討前節的Gaussian核的形狀參數學習規則是否真能正確找出應有的值，在此設計了一個與前節相似的題目，但其形狀不同，其分類公式如下：

$$\begin{aligned}
&\text{當 } (2 \cdot (x+0.5))^2 + (0.5 \cdot (y+0.5))^2 < 0.4 \\
&\text{或 } (0.5 \cdot (x+0.5))^2 + (2 \cdot (y-0.5))^2 < 0.4 \\
&\text{或 } (0.75 \cdot (x-0.5))^2 + (1.5 \cdot (y+0.5))^2 < 0.4 \\
&\text{或 } (1.5 \cdot (x-0.5))^2 + (0.75 \cdot (y-0.5))^2 < 0.4 \\
&\text{則屬第一類，否則為第二類。}
\end{aligned} \tag{37}$$

此問題的範例分佈狀況如圖6所示。

以ARBFN, RBFN, BPN建模的結果如表1所示，可見ARBFN最準確。模型的分類邊界如圖7~圖10所示。



表 1：二維四球分類問題的測試範例誤判率

		BPN一層	BPN二層	RBFN	ARBFN
1	例題1. 基準例題	0.107	0.082	0.044	0.045
2	例題2. 不同形心	0.057	0.047	0.053	0.044
3	例題3. 不同半徑	0.047	0.024	0.044	0.041
4	例題4. 不同形狀	0.150	0.073	0.111	0.039
	平均	0.090	0.057	0.063	0.042

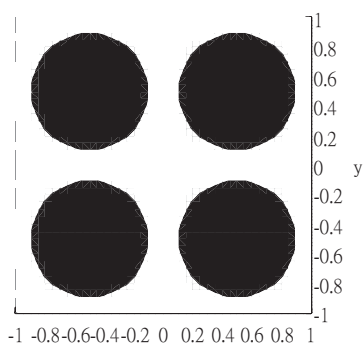


圖 3：例題1.基準例題應有的分類邊界

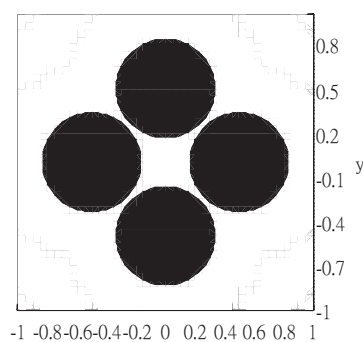


圖 4：例題2.不同形心應有的分類邊界

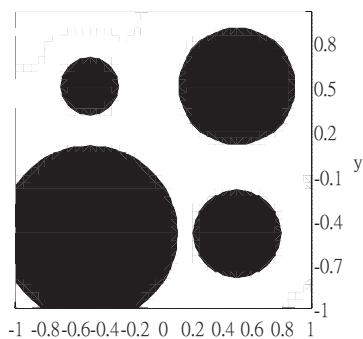


圖 5：例題3.不同半徑應有的分類邊界

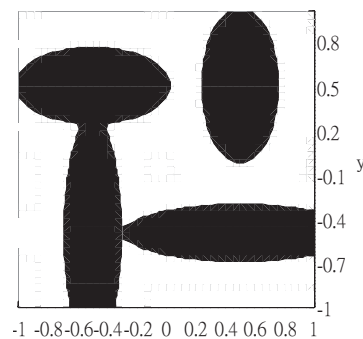
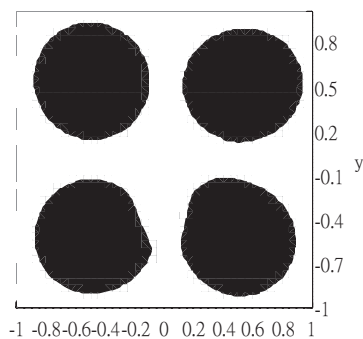
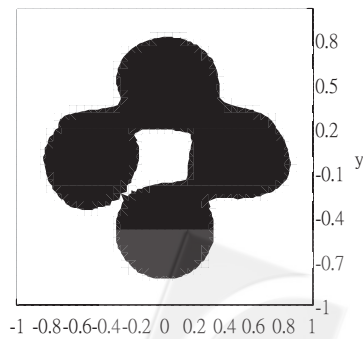


圖 6：例題4.不同形狀應有的分類邊界

圖 7：例題1.基準例題學習的分類邊界
(形狀參數可以自適應調整)圖 8：例題2.不同形心學習的分類邊界
(形狀參數可以自適應調整)

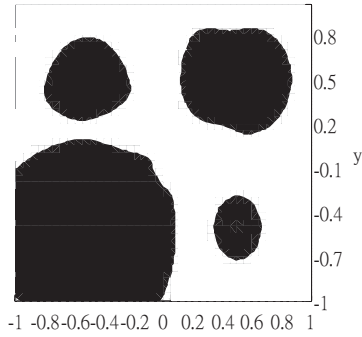


圖 9：例題3.不同半徑學習的分類邊界
(形狀參數可以自適應調整)

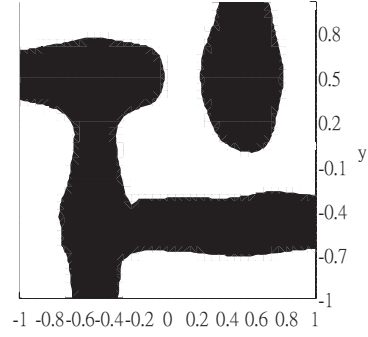


圖10：例題4.不同形狀學習的分類邊界
(形狀參數可以自適應調整)

為了說明以監督式學習推導出的學習規則確實能學得形狀參數值的相對大小，以例題4為例，其Gaussian單元的形心參數、到輸出單元的連結權值、形狀參數如下：

$$C = \begin{bmatrix} -0.52 & -0.47 \\ 0.44 & -0.49 \\ 0.50 & -0.30 \\ 0.51 & 0.50 \\ -0.24 & -0.49 \\ 0.13 & 0.51 \\ -0.44 & 0.51 \\ -0.05 & -0.28 \end{bmatrix} \quad (38)$$

$$W = \begin{bmatrix} 4.4951 \\ 5.2818 \\ -2.9263 \\ 4.8029 \\ -1.2605 \\ -4.1715 \\ 4.3245 \\ -3.6760 \end{bmatrix} \quad (39)$$

$$V = \begin{bmatrix} 2.62 & 0.62 \\ 0.77 & 3.03 \\ 2.53 & 2.60 \\ 3.48 & 1.18 \\ 4.71 & 0.64 \\ 3.93 & 2.75 \\ 1.05 & 3.78 \\ 1.79 & 4.65 \end{bmatrix} \quad (40)$$

由(39)式可知，第1, 2, 4, 7個Gaussian單元到輸出單元的連結權值為正值，代表其Gaussian核代表的是第一類(球)，其對應的形心參數為 $(-0.52, -0.47)$, $(0.44, -0.49)$, $(0.51, 0.50)$, $(-0.44, 0.51)$ 正好與應有的形心 $(-0.50, -0.50)$, $(0.50, -0.50)$, $(0.50, 0.50)$, $(-0.50, 0.50)$ 十分接近；其Gaussian單元的形狀參數為 $(2.62, 0.62)$, $(0.77, 3.03)$, $(3.48, 1.18)$, $(1.05, 3.78)$ ，比例為4.24, 0.26, 2.95, 0.28，與應有的比例4.0, 0.5, 2.0, 0.25相似，可見以監督式學習推導出的學習規則確實能學得形狀參數值的相對大小。

為了進一步證明本文提出的加入形狀參數的演算法優於沒有形狀參數的演算法，在此將原演算法中的形狀參數值在學習過程中全部固定為1.0，其模型的分類邊界如圖11~圖14所示。因為例題1至例題3的分類邊界為圓形，故加入形狀參數的演算法(ARBFN)應該與沒有加入者(RBFN)無太大區別；但在例題4(不同形狀例題)這題上，ARBFN應該要遠優於RBFN才合理。仔細比較圖3~6、圖7~10、圖11~14這三組圖形可以發現：

- (1) RBFN與ARBFN在例題1與例題2產生的邊界並無明顯的優劣之分，這與預期相符。
- (2) RBFN在例題3產生的邊界中，右下方應有圓形未出現；而ARBFN可以產生應有的四個圓形，且形狀亦較正確。這雖與預期不相符，但也顯示ARBFN優於傳統的RBFN，對於具有不同半徑的分類問題有更好的分類能力。
- (3) RBFN在例題4產生的邊界中，其右上方的橢圓橫向寬度大於縱向寬度，與應有的形狀相反；而ARBFN的該橢圓橫向寬度小於縱向寬度，與應有的形狀相同。這顯示ARBFN的「形狀參數」確實發揮了效果，對於具有不同形狀的分類問題有更好的分類能力。

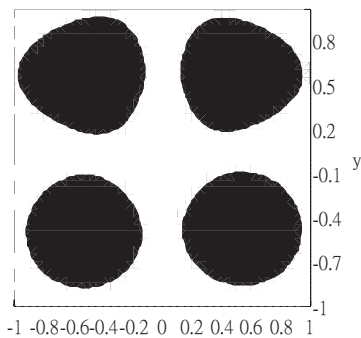


圖11：例題1.基準例題學習的分類邊界
(形狀參數固定為1.0)

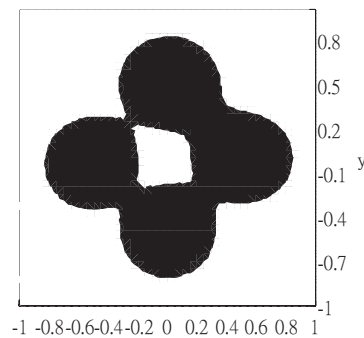


圖12：例題2.不同形心學習的分類邊界
(形狀參數固定為1.0)

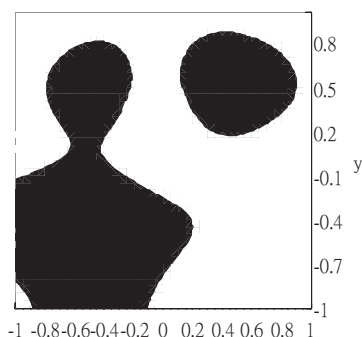


圖13：例題3.不同半徑學習的分類邊界
(形狀參數固定為1.0)

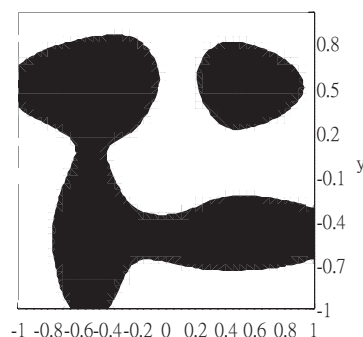


圖14：例題4.不同形狀學習的分類邊界
(形狀參數固定為1.0)

例題5. 二次函數分類問題

為了探討ARBFN的重要性因子是否可以判斷變數的重要性，在此設計了一個分類函數：

$$Y = 1.5(X_1 - 0.5)(X_8 - 0.5) + 1.5(X_4 - 0.5)(X_{10} - 0.5) + (X_5 - 0.5)^2 \quad (41)$$

當 $Y > 0$ 則屬第一類，否則為第二類。

在此函數中， X_1 、 X_8 與 X_4 、 X_{10} 為二組交互作用變數， X_5 為曲率作用變數，其餘 X_2 、 X_3 、 X_6 、 X_7 、 X_9 為無關變數。以ARBFN建模後，依(33)式計算其重要性因子如圖15所示。由圖可知， X_1 、 X_4 、 X_5 、 X_8 、 X_{10} 具有較大的權值，代表它們是較重要的變數，而 X_2 、 X_3 、 X_6 、 X_7 、 X_9 具有較小的權值，代表它們是較不重要的變數。此結論與本例題應有結果一致，證實ARBFN的重要性因子確實可以判斷的輸入變數之重要程度。

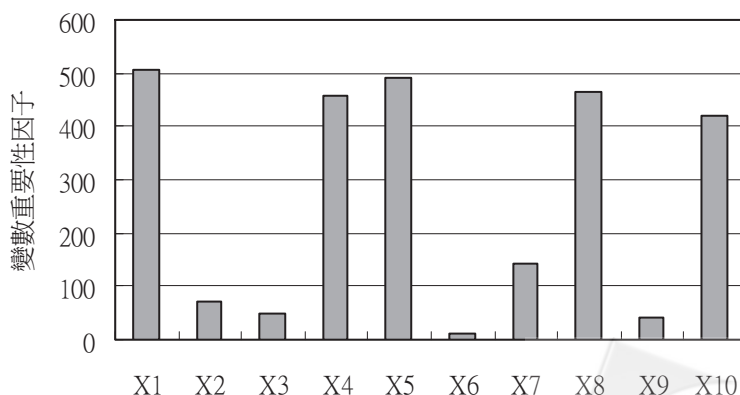


圖15：例題五的輸入變數之重要性因子

肆、實際例題

本節以七個實際應用的題目進行實驗，以比較ARBFN相對於RBFN與BPN的效能，包括休旅車潛在顧客開發(Megaputer Intelligence 2007)、汽車保險潛在顧客開發(The Insurance Company Benchmark 2007)、健身中心會員開發(文少宣 2004)、垃圾郵件過濾(UCI Machine Learning Repository Content Summary 2006)、遙測影像分類(UCI Machine Learning Repository Content Summary 2007)、集集大地震引致山崩(鄒明誠 & 孫志鴻 2005)、森林地表覆蓋分類(Blackard 1998; Forest Cover Type 2007)。

為了檢驗學習的效果，採用五段式交叉驗證(5-fold cross-validation)，即將範例隨機分成五個部份，每次取一個部分做測試範例，其餘四個部分做訓練範例，共作五次，整合所有範例在做為測試範例下的預測結果來評估各方法的準確性。其網路參數設定原則同前節，不再贅述。

以ARBFN, RBFN, BPN建模的結果如表2所示，可見ARBFN最準確。為了比較各法的優劣，在此以單層隱藏層BPN的結果做為基準來進行比較，整理如圖16。為了避免單一例題的影響太大，也將各法在各例題下的優劣給予排名，整理如圖17。由這些圖表可以發現ARBFN最準確，其次依序是二層隱藏層BPN、一層隱藏層BPN，而RBFN最不準確。

表 2：實際例題的測試範例誤判率

	例題名稱	BPN 一層	BPN 二層	RBFN	ARBFN
1	休旅車潛在顧客開發	0.238	0.218	0.212	0.205
2	汽車保險潛在顧客開發	0.340	0.346	0.364	0.364
3	健身中心會員開發	0.328	0.338	0.390	0.312
4	垃圾郵件分類	0.062	0.058	0.083	0.057
5	遙測影像分類	0.120	0.108	0.117	0.102
6	集集大地震引致山崩	0.229	0.214	0.230	0.205
7	森林地表覆蓋分類	0.214	0.208	0.236	0.211
	平均	0.219	0.213	0.233	0.208



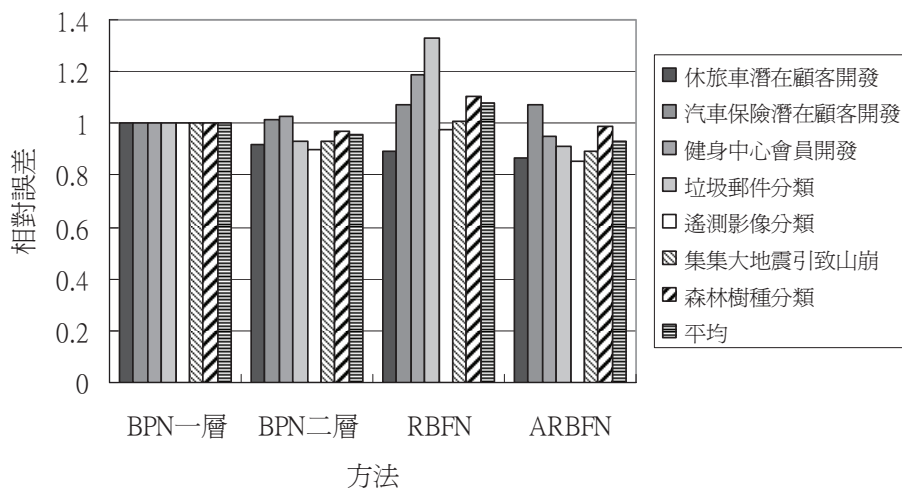


圖 16：實際例題的測試範例誤判率相對大小比較

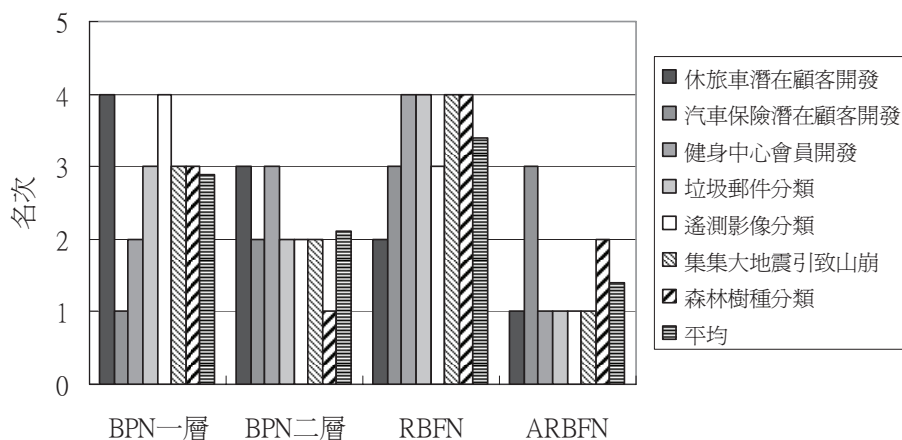


圖 17：實際例題的測試範例誤判率相對名次比較

為了探討ARBFN的重要性因子是否可以判斷變數的重要性，在此以上述的森林地表覆蓋分類實例為例。在此資料集中，森林被分割公尺的格子，其實際森林覆蓋類型是由美國森林服務署(USFS) Region 2資源資訊系統(RIS)資料決定。獨立變數是從美國地質調查署(USGS)與美國森林服務署(USFS)原始資料導出。總共有54個欄位資料，獨立變數包括10個定量連續變數，及44個定性二元變數(4個自然保護區和40種土壤類型)。其中40個用來表示土壤分類類型的二元變數，因其數量龐大，但對分類的影響很小，本研究將其捨去，只取其餘14個變數，如表3所示。原始資料共有58萬(581012)筆資料，有七種覆蓋類型(樹種)。由於各覆蓋類型的資料數目差距極大，本研究採用分層取樣法使各類的資料數目接近，故實際上只用4000筆，各覆蓋類型的資料數目如表4。其中3000筆做為訓練範例，1000筆做為測試範例。

表3：森林地表覆蓋類型實例的輸入變數

變數名稱	資料型態	單位
X1=海拔	連續變數	公尺
X2=方位	連續變數	度
X3=坡度	連續變數	度
X4=對水體水平距離	連續變數	公尺
X5=對水體垂直距離	連續變數	公尺
X6=對道路水平距離	連續變數	公尺
X7=上午九點陰影	連續變數	0 to 255
X8=中午陰影	連續變數	0 to 255
X9=下午三點陰影	連續變數	0 to 255
X10=對火點距離	連續變數	公尺
X11= Rawah Wilderness 荒野區	二元變數	0/1
X12= Neota Wilderness 荒野區	二元變數	0/1
X13= Comanche Peak 荒野區	二元變數	0/1
X14= Cache la Poudre 荒野區	二元變數	0/1

表4：各覆蓋類型的資料數目

編號	覆蓋類型	原始數目	採用數目
1	雲杉木 (Spruce-Fir)	211840	580
2	海灘松樹 (Lodgepole Pine)	283301	557
3	美國黃松木 (Ponderosa Pine)	35754	551
4	楊樹/柳樹 (Cottonwood/Willow)	2747	560
5	白楊樹 (Aspen)	9493	607
6	花旗松 (Douglas-fir)	17367	556
7	矮盤灌叢 (Krummholz)	20510	589
	合計	581012	4000

以ARBFN建模後，依(33)式計算其重要性因子如表5所示。為了相對比較各個輸入變數對各個輸出變數(分類)之重要程度，將表5改以相對大小直條圖表示如圖18。由圖可知，海拔(X_1)重要性明顯高於其他變數，這與樹種分佈受海拔影響的知識相符。此外，水體水平距離(X_4)此一輸入變數對楊樹/柳樹(分類4)的重要性特別高，這與此樹種喜好靠近水體生長的特性相符。可見ARBFN的重要性因子確實可以判斷的輸入變數之重要程度。

表5：森林地表覆蓋分類實例的輸入變數之重要性因子

輸入變數	分類1	分類2	分類3	分類4	分類5	分類6	分類7
X1=海拔	536.4	476.3	306.9	178.5	261.5	264.8	296.8
X2=方位	6.0	3.5	20.2	10.3	5.5	19.3	5.5
X3=坡度	6.2	6.6	4.8	1.3	1.8	5.6	0.9
X4=對水體水平距離	16.6	13.0	65.0	99.3	8.1	36.7	7.5
X5=對水體垂直距離	29.2	28.5	14.7	9.7	6.4	12.4	3.1
X6=對道路水平距離	75.4	137.5	98.1	89.3	142.6	74.5	43.4
X7=上午九點陰影	5.9	8.4	9.9	7.4	8.1	8.7	1.3
X8=中午陰影	17.8	12.8	12.4	6.7	2.9	10.8	5.6
X9=下午三點陰影	4.6	9.0	4.2	5.0	5.0	2.2	4.3
X10=對火點距離	61.4	46.3	118.9	73.0	50.5	95.0	31.8
X11= Rawah Wilderness 荒野區	23.6	33.7	55.0	36.8	17.4	53.0	7.9
X12= Neota Wilderness 荒野區	58.3	59.3	44.7	34.4	33.9	36.2	28.2
X13= Comanche Peak 荒野區	8.5	7.5	11.5	11.9	3.8	8.0	2.5
X14= Cache la Poudre 荒野區	14.6	42.0	41.4	28.8	48.5	31.6	10.4

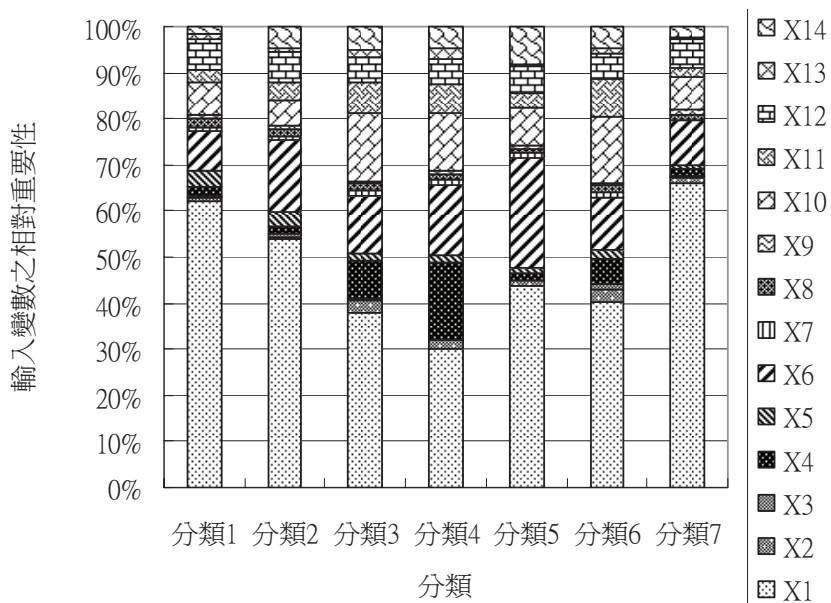


圖18：森林地表覆蓋分類實例的輸入變數之重要性因子之相對大小

伍、結論

徑向基底函數網路(RBFN)常用於分類問題，但它有一個缺點，即視所有自變數有同等地位，故分類邊界是圓形，但事實上每一個自變數對分類的影響力不同，分類邊界是

應該是橢圓形較合理。為克服此一缺點，本文提出具自適應核形狀參數的徑向基底函數網路(ARBFN)，並以監督式學習推導出其學習規則。為證明此一架構優於傳統的徑向基底函數網路，本研究以五個人為的與七個真實的分類例題進行比較。結果歸納如下：

1. 在人為的分類例題中，ARBFN在例題4(不同形狀橢圓分類問題)的表現遠優於RBFN，其原因很可能是ARBFN增加了形狀參數，因此較容易以有限的隱藏單元產生配適資料的分類邊界，而有較小的誤差。
2. 在真實的分類例題中ARBFN優於BPN，但RBFN卻劣於BPN，而ARBFN與RBFN架構的唯一差別是前者增加了形狀參數，可見增加形狀參數可以大幅改善傳統的EBFN架構，使其分類準確度能夠超越BPN。
3. 在人為的與真實的分類例題中，雖然ARBFN並非在每一題中都是最佳的方法，但以整體的誤判率平均值而言，由表1與表2可知，不論是人為的或真實的分類例題，ARBFN都是四個方法中最準確者。且ARBFN在全部11個例題中，其誤判率只有一題微幅高於RBFN，一題等於RBFN，其餘均低於RBFN，可見本文提出的ARBFN確實優於傳統的RBFN。
4. 以監督式學習推導出的學習規則確實能學得形狀參數值的相對大小。
5. ARBFN的重要性因子確實可以判斷的輸入變數之重要程度。

致謝

本研究承蒙國科會贊助(計畫編號96-2221-E-216-032)，特此致謝。

參考文獻

1. 文少宣，2004，類神經網路與決策樹在顧客關係管理應用之比較，中華大學土木工程學系碩士論文。
2. 葉怡成，2006，類神經網路—方法應用與實作，台北：儒林書局。
3. 陳安斌、許育嘉，2004，『整合小波轉換與神經網路於金融投資決策時間序列預測之研究』，資訊管理學報，第十一卷·第一期：139~165頁。
4. 陳安斌、張志良，2001，『基因演算法自動演化之類神經網路在選擇權評價及避險之研究：分析與實證』，資訊管理學報，第七卷·第二期：63~80頁。
5. 鄒明誠、孫志鴻，2004，『預測型模式在空間資料探勘之比較與整合研究』，地理學報，第三十八期：93~109頁。
6. 駱至中、林錦昌，2005，『尋優適應性類神經模糊推論模式於DRGs取巧行為自動檢測之應用』，資訊管理學報，第十二卷·第四期：53~74頁。
7. Berthold, M.R. "The TDRBF: A Shift Invariant Radial Basis Function Network," in *Proc. of the Irish Neural Network Conference*, 1994, pp. 7-12.

8. Blackard, J. A. "Comparison of Neural Networks and Discriminant Analysis in Predicting Forest Cover Types," Ph.D. Dissertation, Department of Forest Sciences, Colorado State University, Fort Collins, Colorado, 1998.
9. Bugmann, G. "Normalized Gaussian Radial Basis Function Networks," *Neurocomputing* (Special Issue on Radial Basis Function Networks) (20:1), 1998, pp. 97-110.
10. Carse, B., Pipe, A.G., Fogarty, T.C. and Hill, T. "Evolving Radial Basis Function Neural Networks Using a Genetic Algorithm," *IEEE International Conference on Evolutionary Computation*, Perth, WA, Australia, 1995, pp. 300-305.
11. Cheng, Y.M. "Adaptive Rival Penalized Competitive Learning and Combined Linear Predictor Model for Financial Forecast and Investment," *International Journal of Neural systems* (8:5), 1997, pp. 517 -534.
12. Forest Cover Type, 2007, (<http://kdd.ics.uci.edu/databases/coverttype/coverttype.html>).
13. Gao, D. and Yang, G. "Adaptive RBF Neural Networks for Pattern Classifications," ?Proceedings of the 2002 International Joint Conference on Neural Networks, 2002, pp. 846-851.
14. Gomm, J.B. and Ding, L.Y. "Selecting Radial Basis Function Network Centers with Recursive Orthogonal Least Squares Training," *IEEE Transactions Neural Networks* (11:2), 2000, pp. 306-314.
15. Han, M. and Xi, J. "Radial Basis Perception Network and Its Applications for Pattern Recognition," *International Joint Conference on Neural Networks*, Honolulu, HI, 2002, pp. 669-674.
16. Megaputer Intelligence, Inc., PolyAnalyst Case Studies, 2007, (<http://www.megaputer.com/>).
17. Moody, J. and Darken, C. "Fast Learning in Networks of Locally-Tuned Processing Units," *Neural Computation* (1:2), 1989, pp. 281-294.
18. Park, J., Harley, R.G., and Venayagamoorthy, G.K. "Indirect adaptive control for synchronous Generator: comparison of MLP/RBF neural networks approach with Lyapunov stability analysis," *IEEE Transactions on Neural Networks* (15:2), 2004, pp. 460- 464.
19. Shibata, K. and Ito, K. "Gauss-Sigmoid Neural Network," *International Joint Conference on Neural Networks*, Washington, DC, 1999, pp. 1203-1208.
20. The Insurance Company Benchmark (COIL 2000), 2007, <http://kdd.ics.uci.edu/databases/tic/tic.html>.
21. UCI Machine Learning Repository Content Summary, Spambase Database, 2006, (<http://www.ics.uci.edu/~mlearn/MLSummary.html>).
22. UCI Machine Learning Repository Content Summary, Statlog Project Databases, 2007, (<http://www.ics.uci.edu/~mlearn/MLSummary.html>).
23. Webb, A.R. and Shannon, S. "Shape-Adaptive Radial Basis Functions," *IEEE Transactions on Neural Networks* (9:6), 1998, pp. 1155-1166.

24. Whitehead, B. A. and Choate, T. D. "Cooperative-Competitive Genetic Evolution of Radial Basis Function Centers and Widths for Time Series Prediction," *IEEE Transactions on Neural Networks* (7:4), 1996, pp. 869-880.
25. Xu, L. "Rival Penalized Competitive Learning for Clustering Analysis RBF Net, and Curve Detection," *IEEE Transactions on Neural Networks* (4:4), 1993, pp. 636 -648.
26. Yeh, I-Cheng "Modeling Chaotic Two-Dimensional Mapping with Fuzzy-Neuron Networks," *Fuzzy Sets and Systems* (105:3), 1999, pp. 421-427.

符號對照表

符號 意義

C	Gaussian單元的核的中心值；
E	誤差函數(能量函數)，用來表示學習的品質；
f	轉換函數；
f'	轉換函數的微分；
H	隱藏層輸出值；
net	處理單元的淨值；
N	處理單元的數目；
Q	Gaussian單元的核的半徑的倒數；
T	輸出層目標輸出值；
V	Gaussian單元的輸入變數的加權值，代表輸入變數的重要程度；
W	Gaussian單元到輸出單元的連結權值；
X	輸入值；
Y	輸出層的推論輸出值；
δ	處理單元的差距量；
$\Delta(\bullet)$	網路權值的修正量；
η	學習速率，控制每次加權值修改的幅度；

下標符號

符號 意義

i	輸入層的處理單元；
j	輸出層的處理單元；
k	隱藏層的處理單元；

