

行政院國家科學委員會專題研究計畫 成果報告

類神經網路演算法之改良 研究成果報告(精簡版)

計畫類別：個別型
計畫編號：NSC 98-2221-E-216-049-
執行期間：98年08月01日至99年07月31日
執行單位：中華大學資訊管理學系

計畫主持人：葉怡成

計畫參與人員：碩士班研究生-兼任助理人員：黃凱鴻
碩士班研究生-兼任助理人員：匡麗麗
碩士班研究生-兼任助理人員：劉佩玲

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中華民國 99 年 09 月 30 日

行政院國家科學委員會補助專題研究計畫成果報告

類神經網路演算法之改良

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC 98-2221-E-216-049-

執行期間：98年8月1日至99年7月31日

執行機構及系所：中華大學資訊管理學系

計畫主持人：葉怡成

計畫參與人員：匡麗麗、高慶恩、黃凱鴻、劉佩玲

成果報告類型(依經費核定清單規定繳交)：精簡報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：

赴國外出差或研習心得報告

赴大陸地區出差或研習心得報告

出席國際學術會議心得報告

國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權，一年二年後可公開查詢

中 華 民 國 99 年 9 月 10 日

摘要

本研究提出最小風險神經網路(Minimum Risk Neural Networks, MRNN)，它以倒傳遞神經網路(BPN)為基礎，加入結構風險最小原理的分類間隔最大化的觀念，其目的為提高 BPN 分類模型的普遍性，克服過度學習，以提高對驗證範例的準確度。為了證明此網路的性能，本研究以 15 個實際應用的分類問題來做測試，並與倒傳遞網路做比較。結果證明最小風險神經網路的模型準確度優於倒傳遞網路。本文並比較 MRNN 與支援向量機的關係，為兩者建立一個統一的理論架構，並證明權值衰減技術是 MRNN 的簡化，為此技術找到較嚴謹的理論基礎。

關鍵字：倒傳遞網路、結構風險最小原理、支援向量機、權值衰減。

Abstract

To enhance the generalization of neural network model, we proposed a novel neural network, Minimum Risk Neural Networks (MRNN), whose principle is the combination of minimizing the sum of squares of error and maximizing the classification margin, based on the principle of structural risk minimization. Therefore, the objective function of MRNN is the combination of the sum of squared error and the sum of squares of the slopes of the classification function. Besides, we derived a more sophisticated formula similar to the traditional weight decay technique from the MRNN, establishing a more rigorous theoretical basis for the technique. This study employed 15 real application examples to test the MRNN. The results led to the following conclusions. (1) As long as the penalty coefficient was in the appropriate range, MRNN performed better than pure MLP. (2) MRNN may perform better in difficult classification problems than MLP using weight decay technique.

Key words: back propagation network, structural risk minimization, SVM, weight decay.

壹、前言

倒傳遞神經網路(back-propagation neural networks, BPN)藉由學習規則來修

正本身的加權值和偏權值，可以建構準確的輸入變數與輸出變數之間的非線性模型，因此應用十分廣泛，例如診斷與預測等領域。因為類神經網路學習的目的在於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 [1,2]：

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

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

網路的連結權值修正量可用最陡坡降法得到 [1,2]

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} = \eta \delta_j X_i \quad (2)$$

其中 W_{ij} = 第 i 個單元與第 j 個單元間的連結加權值； η = 學習速率 (learning rate)，控制每次加權值修改的步幅； δ_j = 第 j 個單元的誤差量； X_i = 第 i 個單元的輸入值。

雖然最小化(1)式的能量函數可以使類神經網路具有建立對訓練範例而言非常精確的非線性模型的能力，即模型具有重現性，但此一模型未必具有預測測試範例的能力，即不具普遍性。這種現象稱之為過度學習(overlearning)。為了克服過度學習，有學者提出權值衰減(weight decay)的技術，即在能量函數中加入權值的平方和[3-7]：

$$E = E_1 + E_2 = \frac{1}{2} \sum_j (T_j - Y_j)^2 + \frac{\lambda}{2} \sum_{k=1}^{N_w} W_k^2 \quad (3)$$

其中 W_k = 第 k 個連結權值； N_w = 網路的連結權值總數； λ = 連結權值懲罰係數，控制權值衰減的程度，其值大於等於 0。

此時網路的連結權值修正量為

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} = \eta (\delta_j X_i - \lambda \cdot W_{ij}) \quad (4)$$

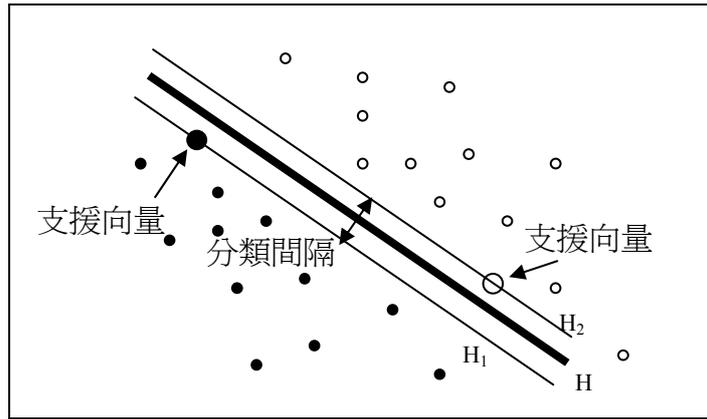
(4)式蘊藏「權值的修正量與權值大小反向」的概念，即當權值為正值時，連結權值修正量會含一個與權值大小成比例的負項；當權值為負值時，連結權值修正量會含一個與權值大小成比例的正項；兩者都有避免權值往極端方向發展的效果，因此這種技術被稱為權值衰減。這個方法不是建立在嚴謹的理論根據上，而是基於啓發式的直覺：「在降低網路輸出單元目標輸出值與推論輸出值之誤差平方和的同時，盡量維持較小之權值平方和，可以產生較具普遍性的網路。」此方法雖然有時可以改善模型的普遍性，但也經常不具實效[3-7]。

支援向量機（Support Vector Machine, SVM）是 Vapnik 等人根據統計學理論提出的一種新的通用學習方法，它是建立在統計學理論的 VC 維（Vapnik Chervonenks Dimension）理論和結構風險最小原理(Structural Risk Minimization Inductive Principle) 的基礎上，能較好地解決小樣本、高維次、非線性和局部極小點等實際問題，已成為機器學習界的研究熱點之一，並成功的應用於分類、函數逼近和時間序列預測等方面[8-14]。

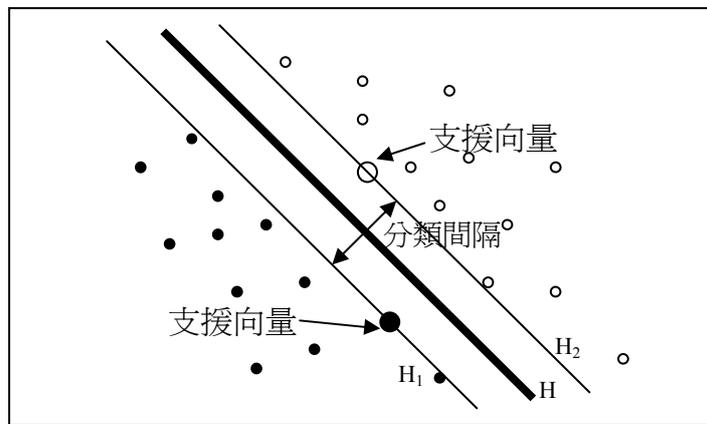
支援向量機是從線性可分情況下的最優分類面發展而來的，其基本思想可用圖 1 的二維情況說明。圖中黑點和白點代表兩類樣本，H 為分類超平面， H_1 、 H_2 分別為過各類中距離分類超平面最近的樣本且平行於分類超平面的平面，它們之間的距離叫做分類間隔（margin）。所謂最優分類面就是要求分類面不但能將兩類正確分開（訓練錯誤率為 0），而且使分類間隔最大，從而提高分類預測能力 [8-14]。在圖 1 中，雖然圖(a)與圖(b)均可完全區分兩類樣本，但圖(b)顯然具有較佳的普遍性。因此在 SVM 中採取下列的能量函數[8-14]：

$$E = E_1 + E_2 = C \sum_{i=1}^k \xi_i + \frac{1}{p(w, b)} \quad (5)$$

其中 C=懲罰係數， $C \geq 0$ ，C 越大，對分類的誤差的懲罰越重。 ξ_i =鬆馳變數， $\xi_i \geq 0$ ，代表第 i 個樣本的分類誤差程度。 $p(w, b)$ =分類間隔大小。



(a)較差的分類超平面(分類間隔較小)



(b)較佳的分類超平面(分類間隔較大)

圖 1 線性可分情況下的最優分類面

在(5)式的能量函數中，第一項是使超平面分類的誤差儘量小，使模型具有重現性；第二項是使樣本到超平面的距離儘量大，從而提高模型分類預測能力，即普遍性。比較(3)式與(5)式知，(5)式的第一項相當於(3)式的誤差平方和項，第二項相當於(3)式的權值平方和項。這個觀察啟發了本研究利用結構風險最小原理推導權值衰減技術理論基礎的動機。

本研究試圖提升類神經網路的普遍性，其原理是以倒傳遞神經網路為基礎，加入結構風險最小原理的分類間隔最大化的觀念。例如圖 2 的一維分類問題，自變數軸的右端樣本屬於 Class A，其分類函數值為 1；自變數軸的左端樣本屬於 Class B，其分類函數值為 0。假設類神經網路經過不同的學習歷程去學這兩個分類的樣本後，產生 H 、 H_1 、 H_2 、 H_3 等分類函數，由圖可知這四個分類函數如以函數值大於 0.5 為界，都可以完全正確地區隔這兩個分類，但 H_1 很明顯有過於

傾向預測 Class A 的可能；H₃ 有過於傾向預測 Class B 的可能；兩者在預測測試範例時都必須冒著較大的誤判風險。H 與 H₂ 分類函數雖然都沒有過於傾向預測 Class A 或 Class B 的可能，且兩者在預測測試範例時都有較小的誤判風險，但由於這兩個分類中間有相當寬闊的區段沒有樣本，因此要產生 H₂ 分類函數是困難的；但如果限制分類函數必須有「最小坡度」，則可能產生 H 分類函數。

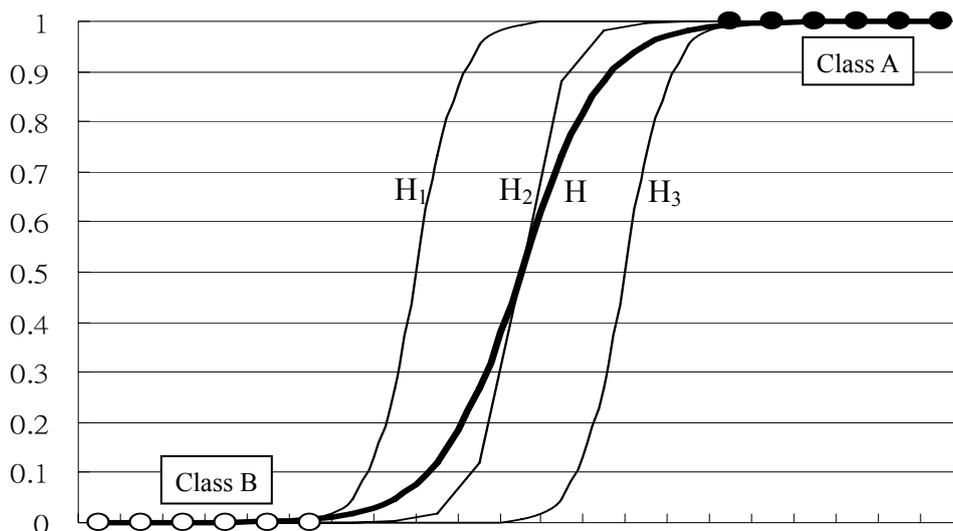


圖 2 一維分類問題與分類函數

因此本文提出「最小風險神經網路」(Minimum Risk Neural Networks, MRNN)，其能量函數除誤差平方和以外，還考慮分類函數的坡度平方和，即一次微分的平方和：

$$E = E_1 + \gamma \cdot E_2 = \frac{1}{2} \sum_j (T_j - Y_j)^2 + \frac{\gamma}{2} \sum_i \sum_j \left(\frac{\partial Y_j}{\partial X_i} \right)^2 \quad (6)$$

其中 γ = 坡度平方和的懲罰係數，控制坡度平方和在能量函數中的比例，其值大於等於 0。

本文將證明(6)式的第二項可以得到與傳統權值衰減相似但更精緻的公式，為權值衰減技術找到嚴謹的理論基礎。倒傳遞神經網路 (BPN)、支援向量機 (SVM)、最小風險神經網路(MRNN)這三種模式之能量函數的比較如表 1 所示。

表 1 三種模式之能量函數的比較

| 模式 | 最小化誤差項 E_1 | 強制普遍化項 E_2 | 強制普遍化原理 |
|-------------------|------------------------------------|---|------------------|
| 權值衰減倒傳遞神經網路 (BPN) | $\frac{1}{2} \sum_j (T_j - Y_j)^2$ | $\frac{\lambda}{2} \sum_{k=1}^{N_w} W_k^2$ | 最小化連結權值的平方和 |
| 支援向量機 (SVM) | $C \sum_{i=1}^k \xi_i$ | $\frac{1}{p(w, b)}$ | 最小化分類間隔的倒數 |
| 最小風險神經網路 (MRNN) | $\frac{1}{2} \sum_j (T_j - Y_j)^2$ | $\frac{\gamma}{2} \sum_i \sum_j \left(\frac{\partial Y_j}{\partial X_i} \right)^2$ | 最小化 Y 對 X 斜率的平方和 |

貳、理論推導

傳統的倒傳遞網路的隱藏層的輸出值公式如下：

$$H_k = f(net_k) = \frac{1}{1 + \exp(-net_k)} \quad (7)$$

$$net_k = \sum_i W_{ik} X_i - \theta_k \quad (8)$$

其中 H_k = 第 k 個隱藏單元的輸入值； net_k = 輸入值之加權乘積和； X_i = 第 i 個輸入單元的輸入值； W_{ik} = 第 i 個輸入單元與第 k 個隱藏單元間的連結加權值； θ_k = 第 k 個隱藏單元的門限值。

輸出層的輸出值公式與傳統的倒傳遞網路相同：

$$Y_j = f(net_j) = \frac{1}{1 + \exp(-net_j)} \quad (9)$$

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

其中 net_j = 隱藏值之加權乘積和； W_{kj} = 第 k 個隱藏單元與第 j 個輸出單元間的連結加權值； θ_j = 第 j 個輸出單元的門限值。

最小風險神經網路的能量函數如(6)式所示，根據最陡坡降法，網路的連結權值修正量可用能量函數對連結權值的偏微分得到

$$\Delta W_{ij} = -\eta \cdot \frac{\partial E}{\partial W_{ij}} = -\eta \cdot \left(\frac{\partial E_1}{\partial W_{ij}} + \gamma \cdot \frac{\partial E_2}{\partial W_{ij}} \right) \quad (11)$$

以下分成二部份推導具有一層隱藏層的神經網路其加權值與偏權值的修正量公式：

● 隱藏層與輸出層間之連結加權值與偏權值的修正量公式

$$-\frac{\partial E_1}{\partial W_{kj}} = -\frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial net_j} \frac{\partial net_j}{\partial W_{kj}} = (T_j - Y_j) \cdot f'(net_j) \cdot H_k \quad (12)$$

$$\text{令 } \delta_j \equiv (T_j - Y_j) \cdot f'(net_j) \quad (13)$$

則

$$-\frac{\partial E_1}{\partial W_{kj}} = \delta_j \cdot H_k \quad (14)$$

$$-\frac{\partial E_2}{\partial W_{kj}} = -\frac{\partial}{\partial W_{kj}} \left(\frac{1}{2} \sum_i \sum_j \left(\frac{\partial Y_j}{\partial X_i} \right)^2 \right) \quad (15)$$

上式中的一階微分可用連鎖律推導

$$\frac{\partial Y_j}{\partial X_i} = \sum_l \frac{\partial Y_j}{\partial net_j} \frac{\partial net_j}{\partial H_l} \frac{\partial H_l}{\partial net_l} \frac{\partial net_l}{\partial X_i} = \sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \quad (16)$$

將(16)式代入(15)式得

$$\begin{aligned} -\frac{\partial E_2}{\partial W_{kj}} &= -\frac{\partial}{\partial W_{kj}} \left(\frac{1}{2} \sum_i \sum_j \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right)^2 \right) \\ &= -\frac{1}{2} \sum_i \frac{\partial}{\partial W_{kj}} \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right)^2 \\ &= -\sum_i \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right) \left(f'(net_j) \cdot f'(net_k) \cdot W_{ik} \right) \end{aligned} \quad (17)$$

故

$$\begin{aligned} \Delta W_{kj} &= -\eta \cdot \left(\frac{\partial E_1}{\partial W_{kj}} + \gamma \cdot \frac{\partial E_2}{\partial W_{kj}} \right) \\ &= \eta \cdot \left(\delta_j H_k - \gamma \cdot \sum_i \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right) \left(f'(net_j) \cdot f'(net_k) \cdot W_{ik} \right) \right) \end{aligned} \quad (18)$$

由於在上式中，第二項（即參數 γ 所乘的項）之大小與輸入單元及隱藏單元

的數目成比例，故爲了使第二項與第一項 $\delta_j H_k$ 能夠平衡，在實際計算 ΔW_{kj} 時，第二項須除以 $\sqrt{N_{inp} N_{hid}}$ ，其中 N_{inp} = 輸入單元數目， N_{hid} = 隱藏單元數目。

● 輸入層與隱藏層間之連結加權值與偏權值的修正量公式

$$\begin{aligned}
 -\frac{\partial E_1}{\partial W_{ik}} &= -\frac{\partial E}{\partial H_k} \frac{\partial H_k}{\partial net_k} \frac{\partial net_k}{\partial W_{ik}} \\
 &= -\left(\sum_j \frac{\partial E}{\partial Y_j} \frac{\partial Y_j}{\partial net_j} \frac{\partial net_j}{\partial H_k} \right) \cdot f'(net_k) \cdot X_i \\
 &= -\left(\sum_j -(T_j - Y_j) \cdot f'(net_j) \cdot W_{kj} \right) \cdot f'(net_k) \cdot X_i \\
 &= \left(\sum_j \delta_j W_{kj} \right) \cdot f'(net_k) \cdot X_i \tag{19}
 \end{aligned}$$

$$\text{令 } \delta_k \equiv \left(\sum_j \delta_j W_{kj} \right) \cdot f'(net_k) \tag{20}$$

則

$$-\frac{\partial E_1}{\partial W_{ik}} = \delta_k \cdot X_i \tag{21}$$

$$-\frac{\partial E_2}{\partial W_{ik}} = -\frac{\partial}{\partial W_{ik}} \left(\frac{1}{2} \sum_i \sum_j \left(\frac{\partial Y_j}{\partial X_i} \right)^2 \right) \tag{22}$$

將(16)式代入(22)式得

$$\begin{aligned}
 -\frac{\partial E_2}{\partial W_{ik}} &= -\frac{\partial}{\partial W_{ik}} \left(\frac{1}{2} \sum_i \sum_j \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right)^2 \right) \\
 &= -\frac{1}{2} \sum_j \frac{\partial}{\partial W_{ik}} \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right)^2 \\
 &= -\sum_j \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right) \left(f'(net_j) \cdot W_{kj} \cdot f'(net_k) \right) \tag{23}
 \end{aligned}$$

故

$$\begin{aligned}
 \Delta W_{ik} &= -\eta \cdot \left(\frac{\partial E_1}{\partial W_{ik}} + \gamma \cdot \frac{\partial E_2}{\partial W_{ik}} \right) \\
 &= \eta \cdot \left(\delta_k X_i - \gamma \cdot \sum_j \left(\sum_l f'(net_j) \cdot W_{lj} \cdot f'(net_l) \cdot W_{il} \right) \left(f'(net_j) \cdot W_{kj} \cdot f'(net_k) \right) \right)
 \end{aligned}$$

(24)

由於在上式中，第二項 (即參數 γ 所乘的項) 之大小與輸出單元及隱藏單元的數目成比例，故為了使第二項與第一項 $\delta_k X_i$ 能夠平衡，在實際計算 ΔW_{ik} 時，第二項須除以 $\sqrt{N_{out} N_{hid}}$ ，其中 N_{out} =輸出單元數目， N_{hid} =隱藏單元數目。

● 最小風險神經網路與權值衰減技術之關係

為尋找最小風險神經網路與權值衰減技術之關係，因(18)式中的 $f'(net_j)$ 與 $f'(net_k)$ 必為正值，並不影響(18)式的正負號，而因此可以考慮忽略之，故可簡化得

$$\Delta W_{kj} = \eta \cdot \left(\delta_j H_k - \gamma \cdot \sum_i \left(\sum_l W_{il} \cdot W_{lj} \right) \cdot W_{ik} \right) \quad (25)$$

同理，(24)式可簡化為

$$\Delta W_{ik} = \eta \cdot \left(\delta_k X_i - \gamma \cdot \sum_j \left(\sum_l W_{il} \cdot W_{lj} \right) \cdot W_{kj} \right) \quad (26)$$

比較(25)式與(26)式和權值衰減技術之(4)式可知，兩者的基本原則類似，都蘊藏「權值的修正量與權值大小反向」的概念，因此傳統的權值衰減技術可視為最小風險神經網路的簡化，為權值衰減技術找到嚴謹的理論基礎。

參、應用實例

為了證明 MRNN 可以應用在實際分類應用上，本研究選擇山崩分類問題[15]來進行測試，並與 BPN 做比較。台灣由於受地理與氣候影響，水土災害一直是主要的天然災害。再加上多處山坡地過度開發，往往在地震發生時帶來嚴重的山崩及土石鬆動，造成嚴重的土石崩塌災害發生。1999 年 9 月 21 日於台灣中部發生芮氏規模 7.3 的地震，而此次地震亦獲得大量的坡地破壞資料及地震紀錄，可供學者們進行地震對山崩影響的研究。

本例題資料來自文獻[15]所提供的 60000 筆資料。首先隨機抽樣山崩、未山崩各 2000 筆樣本，做為本研究的數據。再隨機取 3200 筆為訓練範例，800 筆樣本為驗證範例。本研究有 15 個輸入變數，如表 2。結果如圖 3。由圖可知，在適當的 γ 參數時($\gamma=0.03\sim 0.3$)，MRNN 誤判率低於傳統的 BPN。當 γ 達最佳時($\gamma=0.1$)，MRNN 誤判率(19.1%)明顯比 BPN(20.1%)低。

此外，本例題也以 SVM [16] 建立分類模型，並以網格法決定最佳懲罰係數

C 與核心係數 g ，其結果如圖 4 所示。由圖表可知，雖然在最佳的懲罰係數與核心係數下，最小誤差可達 19.1%，與 MRNN 相同，但與 MRNN 相較，SVM 有幾個缺點：(1) 其最佳結果是嘗試了 $8 \times 8 = 64$ 種懲罰係數與核心係數組合，故較為耗時；(2) 在使用大量參數調適下，模型的真實普遍化能力可能被高估；(3) 由圖 18 可以看出 SVM 分類模型的準確度對懲罰係數與核心係數相當敏感，只有很少的參數組合其誤判率低於 20% (圖 18 中的白色區域)。

表 2 輸入變數表

| 編 | 變數名稱 | 編 | 變數名稱 |
|---|-----------|----|-------------|
| 1 | 高程 | 9 | 垂直地表加速度 |
| 2 | 坡向 | 10 | 東西向地表加速度 |
| 3 | 坡度 | 11 | 南北向地表加速度 |
| 4 | 距離車籠埔斷層距離 | 12 | 九格點之平均坡度 |
| 5 | 距離斷層破碎帶距離 | 13 | 九格點之最大最小坡度差 |
| 6 | 距離道路距離 | 14 | 九格點之平均坡向 |
| 7 | 距水系距離 | 15 | 九格點之最大最小坡度差 |
| 8 | 距震央距離 | | |

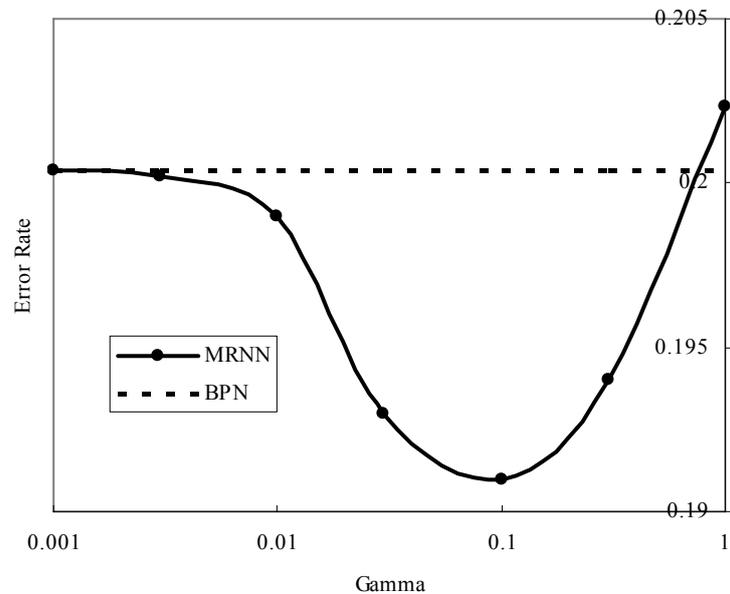


圖 3 應用實例的 γ 參數與誤判率的關係

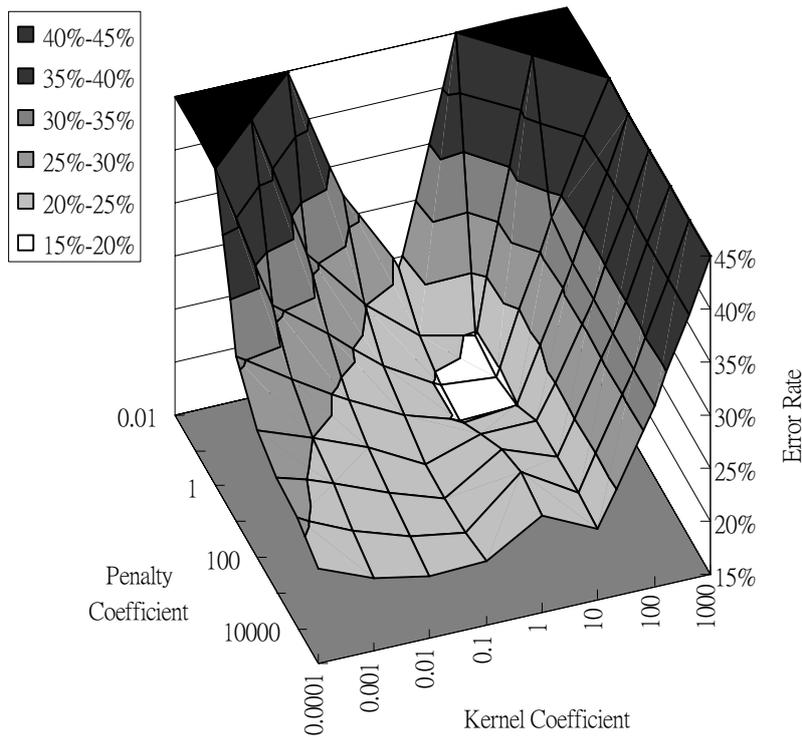


圖 4 應用實例以 SVM 建模之結果(網格法 3D 圖)

爲了驗證 MRNN 在少量樣本數下的效果，在此取原訓練樣本的 1/5、1/10、1/20 的樣本作爲建模的訓練集，其結果如圖 5。由圖可知，在 1/5 的樣本下，MRNN 在適當參數 γ 下，其誤判率明顯比 BPN(圖上虛線)低。但隨著樣本的減少，兩者之間的差異變小。

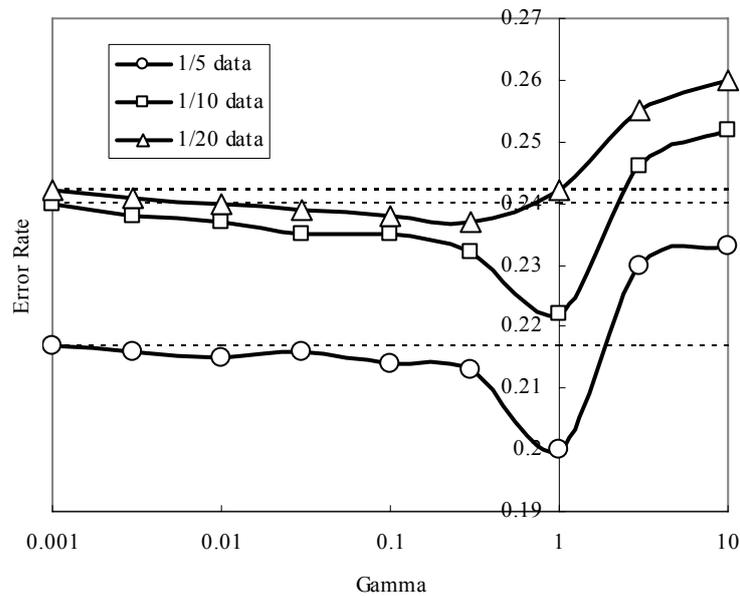


圖 5 應用實例在不同樣本數下 γ 參數與誤判率的關係

我們還嘗試 15 個 UCI 的實際數據資料庫(表 3)比較 MRNN，MLPWD 和支持向量機的性能。爲了評估學習的有效性，我們採用了 10 倍交叉驗證。此外，爲避免初始連接權值的影響，錯誤率是 30 組不同初始連接權值的平均結果。爲評估各種方法之間的差異是否顯著，採用 t 檢定。結果如表 4，顯示 MRNN 的準確性優於權值衰退法，而與 SVM 相近。

Table 3. Descriptions of the 15 UCI data sets

| UCI data sets | Input variable | Classes | data |
|-------------------|----------------|---------|------|
| SPAMBASE | 57 | 2 | 4601 |
| Landsat Satellite | 36 | 6 | 6435 |
| Forest cover | 14 | 7 | 4000 |
| Iris | 4 | 3 | 150 |
| Insurance | 7 | 2 | 700 |
| Glass | 9 | 6 | 214 |
| Shuttle | 9 | 3 | 5000 |
| Vowel | 10 | 11 | 990 |
| Wine | 13 | 3 | 178 |
| Letter | 16 | 26 | 2000 |
| Image | 18 | 7 | 2310 |
| Vehicle | 18 | 4 | 846 |
| German | 19 | 2 | 1000 |
| Heart | 20 | 2 | 270 |
| Thyroid | 21 | 3 | 7200 |

Table 4. Testing results of error rate of the 15 UCI data sets

| UCI data sets | Benchmark | | | MRNN | | MRNN vs. Benchmark t-test (Significance=5%) | |
|---------------|-----------|--------|--------|--------|--------|---|----------|
| | MLPWD | | SVM | | | | |
| | Avg. | Std. | | Avg. | Std. | MLPWD | SVM |
| SPAMBASE | 0.0642 | 0.0027 | 0.0653 | 0.0631 | 0.0018 | 0.037 * | <0.001 * |
| Landsat | 0.0981 | 0.0016 | 0.098 | 0.0974 | 0.0013 | 0.036 * | 0.008 * |
| Forest cover | 0.232 | 0.002 | 0.215 | 0.208 | 0.003 | <0.001 * | <0.001 * |
| Iris | 0.0270 | 0 | 0.027 | 0.0270 | 0 | >0.5 | >0.5 |
| Insurance | 0.3366 | 0.0131 | 0.3365 | 0.3363 | 0.0160 | 0.468 | 0.473 |
| Glass | 0.2675 | 0.0036 | 0.2665 | 0.2667 | 0.0047 | 0.248 | >0.5 |
| Shuttle | 0.0049 | 0.0001 | 0.004 | 0.0040 | 0.0001 | <0.001 * | >0.5 |
| Vowel | 0.4123 | 0.0091 | 0.4052 | 0.3983 | 0.0096 | <0.001 * | <0.001 * |
| Wine | 0.0116 | 0.0002 | 0.0115 | 0.0113 | 0.0002 | <0.001 * | <0.001 * |

| | | | | | | | |
|---------|--------|--------|--------|--------|--------|----------|----------|
| Letter | 0.3474 | 0.0071 | 0.3418 | 0.3315 | 0.0073 | <0.001 * | <0.001 * |
| Image | 0.0422 | 0.0008 | 0.0422 | 0.0421 | 0.0009 | 0.330 | 0.281 |
| Vehicle | 0.1240 | 0.0020 | 0.1232 | 0.1230 | 0.0015 | 0.025 * | 0.294 |
| German | 0.2393 | 0.0071 | 0.2365 | 0.2362 | 0.0052 | 0.034 * | 0.397 |
| Heart | 0.1430 | 0.0019 | 0.143 | 0.1430 | 0.0019 | >0.5 | >0.5 |
| Thyroid | 0.0241 | 0.0002 | 0.0231 | 0.0198 | 0.0002 | <0.001 * | <0.001 * |

肆、結論

本研究提出最小風險神經網路，它以倒傳遞神經網路(BPN)為基礎，加入結構風險最小原理的分類間隔最大化的觀念，其目的為为了提高 BPN 分類模型的普遍性，以提高對驗證範例的準確度。由 15 個應用例題的驗證歸納得到 MRNN 的準確性優於權值衰退法的結論。

本研究與其它文獻不同之處與創新的貢獻簡述如下：

1. 改善BPN的普遍化預測能力

BPN與SVM均可建立複雜的非線性模型。傳統的BPN是建立在誤差平方和最小化的基礎上，雖然相對於SVM而言，BPN具有理論簡明易懂、計算過程簡單的優點，但它也有容易過度學習，普遍化預測能力差，即對驗證範例的準確度低的缺點。而SVM是建立在分類間隔最大化的基礎上，雖然相對於BPN而言，SVM具有普遍化預測能力佳的優點，但它也有理論繁複難懂、計算過程複雜的缺點。本研究提出的MRNN是以BPN依賴的誤差平方和最小化為基礎，加入SVM依賴的分類間隔最大化的觀念，不但提高了BPN的普遍化預測能力，也避免了SVM理論繁複難懂的缺點。

2. 發現SVM與神經網路的數學理論相似架構

SVM是建立在分類間隔最大化的基礎上；而傳統的BPN則建立在誤差平方和最小化的基礎上；兩者看似無關。本研究經由提出MRNN發現了SVM與神經網路的數學理論相似架構，為建立神經網路與SVM的統一理論架構提供一個可能的研究方向。

3. 證明權值衰減技術是MRNN的簡化

權值衰減技術長期以來一直被視為是一個可以提高神經網路普遍化預測能力的方法，但此法是建立在經驗法則的基礎上。本研究提出的MRNN經過簡化後

與權值衰減技術相似但較精緻，因此權值衰減技術可視為MRNN的簡化，而MRNN也可視為權值衰減技術的進階，為權值衰減技術找到較嚴謹的理論基礎。

未來的研究方向包括：

1. 探討決定參數 γ 的自動化方法。
2. 探討 MRNN 在迴歸問題的效果。

參考文獻

1. Haykin, S., *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, NJ (2005).
2. Rumelhart, D.E., Hinton, G.E., and Williams, R.J., "Learning representations by back-propagation errors," *Nature*, Vol. 323, pp. 533-536 (1986).
3. Hinton, G. E. and Camp, D., "Keeping the neural networks simple by minimizing the description length of the weights," *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, pp. 5-13 (1993).
4. Treadgold, N.K. and Gedeon, T.D., "Simulated annealing and weight decay in adaptive learning: the SARPROP algorithm," *IEEE Transactions on Neural Networks*, Vol.9, No.4, pp. 662-668 (1998).
5. Krogh, A. and Hertz, J. A., "A simple weight decay can improve generalization," In J.E. Moody, S.J. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems 4*, pages 450--957, San Mateo, CA (1992).
6. Gupta, A. and Lam, S.M., "Weight decay backpropagation for noisy data," *Neural Networks*, Vol.11, No. 6, pp. 1127-1137 (1998).
7. Ramíreza, M.C.V. Velhob, H.F.C. and Ferreiraa, N.J. "Artificial neural network technique for rainfall forecasting applied to the São Paulo region," *Journal of Hydrology*, Vol. 301, No. 1-4, pp.146-162 (2005).
8. Cortes, F., and Vapnik, V., 1995, "Support vector networks", *Machine Learning*, Vol. 20, No. 3, pp.273-297 (1995).
9. Vapnik V.N., *The Nature of Statistical Learning Theory*, Springer-Verlag: New York (1995).
10. Drucker, H., Wu, D., and Vapink V., "Support vector machines for spam categorization", *IEEE Transactions on Neural Networks*, Vol.10, No. 5, pp.1048-1054 (1999).

11. Burges, C., “A tutorial on support vector machines for pattern recognition”, *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp.121-167 (1998).
12. Fan, R.E., Chen, P.H., and Lin, C.J.. “Working set selection using second order information for training support vector machines,” *The Journal of Machine Learning Research*, Vol. 6, 1889–1918 (2005).
13. Glasmachers, T. and Igel, C., “Second-order SMO improves SVM online and active learning,” *Neural Computation*, Vol. 20, No. 2, pp.374-382 (2008).
14. Glasmachers, T. and Igel, C. “Maximum-gain working set selection for SVMs,” *The Journal of Machine Learning Research*, Vol. 7, pp.1437-1466 (2006).
15. Tsou, M. C. and Sun, C. H. “A Comparative and integrated study of a predictive model in spatial data mining,” *Journal of Geographical Science*, Vol. 38, pp. 93-109 (2004).
16. Chang, C.C. and C.-J. Lin. LIBSVM: a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

計畫成果自評

研究計畫為「提出最小風險神經網路(Minimum Risk Neural Networks, MRNN)，它以倒傳遞神經網路(BPN)為基礎，加入結構風險最小原理的分類間隔最大化的觀念，其目的為为了提高 BPN 分類模型的普遍性，克服過度學習，以提高對驗證範例的準確度。本研究將比較 MRNN 與支援向量機的關係，為兩者建立一個統一的理論架構，並證明權值衰減技術是 MRNN 的簡化，為此技術找到較嚴謹的理論基礎。」

研究內容為「本研究提出最小風險神經網路(Minimum Risk Neural Networks, MRNN)，它以倒傳遞神經網路(BPN)為基礎，加入結構風險最小原理的分類間隔最大化的觀念，其目的為为了提高 BPN 分類模型的普遍性，克服過度學習，以提高對驗證範例的準確度。為了證明此網路的性能，本研究以 15 個實際應用的分類問題來做測試，並與倒傳遞網路做比較。結果證明最小風險神經網路的模型準確度優於倒傳遞網路。本文並比較 MRNN 與支援向量機的關係，為兩者建立一個統一的理論架構，並證明權值衰減技術是 MRNN 的簡化，為此技術找到較嚴謹的理論基礎。」

因此研究內容與原計畫完全相符程度。

本研究的成果已寫成英文論文並投稿，目前正在審稿中。另外，這一年內作者發表了相關著作：

1. Yeh, I-Cheng and Wei-Lun Cheng, “First and second order sensitivity analysis of MLP,” *Neurocomputing*, Vol.73, No.10, pp. 2225–2233 (2010). **(SCI, IF=1.44)**
2. 葉怡成、程韋綸，「可衡量輸入變數重要性的神經網路—灰箱倒傳遞網路」，*先進工程學刊*，第 5 卷，第 1 期，第 41-48 頁(2010)。

因此已達成預期目標。

本研究與其它文獻不同之處與創新的貢獻簡述如下：

1. 改善BPN的普遍化預測能力

BPN與SVM均可建立複雜的非線性模型。傳統的BPN是建立在誤差平方和最小化的基礎上，雖然相對於SVM而言，BPN具有理論簡明易懂、計算過程簡單的優點，但它也有容易過度學習，普遍化預測能力差，即對驗證範例的準確度

低的缺點。而SVM是建立在分類間隔最大化的基礎上，雖然相對於BPN而言，SVM具有普遍化預測能力佳的優點，但它也有理論繁複難懂、計算過程複雜的缺點。本研究提出的MRNN是以BPN依賴的誤差平方和最小化為基礎，加入SVM依賴的分類間隔最大化的觀念，不但提高了BPN的普遍化預測能力，也避免了SVM理論繁複難懂的缺點。

2. 發現SVM與神經網路的數學理論相似架構

SVM是建立在分類間隔最大化的基礎上；而傳統的BPN則建立在誤差平方和最小化的基礎上；兩者看似無關。本研究經由提出MRNN發現了SVM與神經網路的數學理論相似架構，為建立神經網路與SVM的統一理論架構提供一個可能的研究方向。

3. 證明權值衰減技術是MRNN的簡化

權值衰減技術長期以來一直被視為是一個可以提高神經網路普遍化預測能力的方法，但此法是建立在經驗法則的基礎上。本研究提出的MRNN經過簡化後與權值衰減技術相似但較精緻，因此權值衰減技術可視為MRNN的簡化，而MRNN也可視為權值衰減技術的進階，為權值衰減技術找到較嚴謹的理論基礎。

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期：99 年 9 月 10 日

| | | | |
|--------|---|---------|------------------|
| 計畫編號 | NSC-98-2221-E-216-049- | | |
| 計畫名稱 | 類神經網路演算法之改良 | | |
| 出國人員姓名 | 葉怡成 | 服務機構及職稱 | 中華大學 資訊管理系 教授 |
| 會議時間 | 2010/7/11-210/7/14 | 會議地點 | 山東省青島市 |
| 會議名稱 | (中文) 2010 機器學習和控制論國際研討會 (英文) International Conference on Machine Learning and Cybernetics 2010 | | |
| 發表論文題目 | (中文) 複合轉換函數網路 (英文) Hybrid Transfer Function Networks (中文) 可調內核形狀參數的徑向基函數網路 (英文) Radial Basis Function Networks with Adjustable Kernel Shape Parameters (中文) 以遺傳運算樹作混凝土強度建模 (英文) Modeling Concrete Strength Using Genetic Operation Trees | | |

一、參加會議經過

此次 2010 機器學習和控制論國際研討會在 2010/7/11-210/7/14 山東省青島市招開。本人共發表三篇

(1) Hybrid Transfer Function Networks

(2) Radial Basis Function Networks with Adjustable Kernel Shape Parameters

(3) Modeling Concrete Strength Using Genetic Operation Trees

這三篇都是國科會 NSC-98-2221-E-216-049 的相關成果。

其中第一篇 Hybrid Transfer Function Networks 獲得大會入選最

佳論文獎候選(共五篇入選)，並在 2010/7/12 日下午與議場發表。並在 2010/7/13 晚宴中發表得獎名單。最後是由另一位也是來自台灣的學者獲獎。

二、與會心得

這次會議的研究主題包含了許多機器學習的相關議題，並請學者專題演獎，包括：

Multiple Classifier Systems

How to disseminate your research results: essentials of effective publishing

Multiple Kernel Learning and Feature Space Denoising

Incompleteness in Data for Decision Making

其中 Multiple Classifier Systems 似乎是一個重要趨勢，因為已有很多分類器被提出，如何更進一步整合多個分類器是一個提升準確率的可行方法。

三、考察參觀活動(無是項活動者略)

四、建議

青島市是一個風景秀麗的城市，觀光資源豐富，是 2008 奧運帆船項目場地。但除了日韓以外，外國觀光客不多。還有可以改進之處。不過其金融中心的建設相當集中，潛力不可忽視。

五、攜回資料名稱及內容

(1) 會議論文光碟一份

(2) 議程文件一份

無研發成果推廣資料

98 年度專題研究計畫研究成果彙整表

| 計畫主持人：葉怡成 | | 計畫編號：98-2221-E-216-049- | | | | |
|------------------|-----------------|-------------------------|-----------------|------------|------|--|
| 計畫名稱：類神經網路演算法之改良 | | | | | | |
| 成果項目 | | 量化 | | | 單位 | 備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等） |
| | | 實際已達成數（被接受或已發表） | 預期總達成數（含實際已達成數） | 本計畫實際貢獻百分比 | | |
| 國內 | 論文著作 | 期刊論文 | 1 | 1 | 100% | 篇 葉怡成、程韋綸(2010),「可衡量輸入變數重要性的神經網路—灰箱倒傳遞網路」,先進工程學刊,第5卷,第1期,第41-48頁。 |
| | | 研究報告/技術報告 | 0 | 0 | 100% | |
| | | 研討會論文 | 0 | 0 | 100% | |
| | | 專書 | 0 | 0 | 100% | |
| | 專利 | 申請中件數 | 0 | 0 | 100% | 件 |
| | | 已獲得件數 | 0 | 0 | 100% | |
| | 技術移轉 | 件數 | 0 | 0 | 100% | 件 |
| | | 權利金 | 0 | 0 | 100% | 千元 |
| | 參與計畫人力 (本國籍) | 碩士生 | 4 | 4 | 100% | 人次 |
| | | 博士生 | 0 | 0 | 100% | |
| 博士後研究員 | | 0 | 0 | 100% | | |
| 專任助理 | | 0 | 0 | 100% | | |
| 國外 | 論文著作 | 期刊論文 | 1 | 1 | 100% | 篇 I-Cheng Yeh and Wei-Lun Cheng (2010), 'First and Second Order Sensitivity Analysis of MLP,' Neurocomputing, Vol. 73, No. 10, pp. 2225 - 2233. (SCI) |
| | | 研究報告/技術報告 | 0 | 0 | 100% | |

| | | | | | | | |
|-----------------|--|--------|---|---|------|-----|--|
| | | | | | | | (1) I-CHENG YEH, XIN-YING ZHANG, CHONG WU, KUAN-CHIEH HUANG (2010). ' RADIAL BASIS FUNCTION NETWORKS WITH ADJUSTABLE KERNEL SHAPE PARAMETERS,' 2010 INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND CYBERNETICS (ICMLC), 1482-1485, QINGDAO, CHINA. (EI) |
| | | 研討會論文 | 2 | 1 | 100% | | (2) I-CHENG YEH, CHUNG-CHIH CHEN, KUAN-CHIEH HUANG (2010). ' HYBRID TRANSFER FUNCTION NETWORKS,' 2010 INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND CYBERNETICS (ICMLC), 37-42, QINGDAO, CHINA. (nominated as candidate for the Lotfi Zadeh Best Paper Awards) (EI) |
| | | 專書 | 0 | 0 | 100% | 章/本 | |
| 專利 | | 申請中件數 | 0 | 0 | 100% | 件 | |
| | | 已獲得件數 | 0 | 0 | 100% | | |
| 技術移轉 | | 件數 | 0 | 0 | 100% | 件 | |
| | | 權利金 | 0 | 0 | 100% | 千元 | |
| 參與計畫人力 (外國籍) | | 碩士生 | 0 | 0 | 100% | 人次 | |
| | | 博士生 | 0 | 0 | 100% | | |
| | | 博士後研究員 | 0 | 0 | 100% | | |
| | | 專任助理 | 0 | 0 | 100% | | |

| | |
|---|---|
| <p style="text-align: center;">其他成果</p> <p>(無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)</p> | 無 |
|---|---|

| | 成果項目 | 量化 | 名稱或內容性質簡述 |
|---|-----------------|----|-----------|
| 科 教 處 計 畫 加 填 項 目 | 測驗工具(含質性與量性) | 0 | |
| | 課程/模組 | 0 | |
| | 電腦及網路系統或工具 | 0 | |
| | 教材 | 0 | |
| | 舉辦之活動/競賽 | 0 | |
| | 研討會/工作坊 | 0 | |
| | 電子報、網站 | 0 | |
| | 計畫成果推廣之參與(閱聽)人數 | 0 | |

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以 100 字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以 100 字為限）

國內期刊論文一篇

國外期刊論文一篇(SCI)

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

本研究提出最小風險神經網路(Minimum Risk Neural Networks, MRNN)，它以倒傳遞神經網路(BPN)為基礎，加入結構風險最小原理的分類間隔最大化的觀念，其目的為了提高 BPN 分類模型的普遍性，克服過度學習，以提高對驗證範例的準確度。為了證明此網路的性能，本研究以 12 個實際應用的分類問題來做測試，並與倒傳遞網路做比較。結果證明最小風險神經網路的模型準確度優於倒傳遞網路。本文並比較 MRNN 與支援向量機的關係，為兩者建立一個統一的理論架構，並證明權值衰減技術是 MRNN 的簡化，為此技術找到較嚴謹的理論基礎。