橢圓空間機率神經網路

葉怡成
中華大學資訊管理學系

林冠呈
中華大學資訊管理學系

摘要

本研究提出橢圓空間機率神經網路（Ellipse-Space Probabilistic Neural Networks, EPNN），它擁有三種可透過訓練來修正的網路參數：代表輸入變數重要性的變數權值、代表樣本有效範圍的核寬倒數、及代表樣本可靠程度的資料權值。這些網路參數可以提升模型的準確度，並計算出重要性指標，以提供評估輸入變數重要性的能力。為證明此網路的性能，本研究以三個人為的分類問題以及七個實際的分類問題來做測試，並與倒傳遞網路及機率神經網路做比較。結果證明：（1）在人為的分類問題EPNN的模型準確度只略低於BPN，而遠高於PNN；在實際的分類問題EPNN的模型準確度明顯高於BPN與PNN。（2）重要性指標確實可以顯示輸入變數對輸出變數的重要程度，使模型具有解釋能力。

關鍵字：類神經網路、機率神經網路、變數重要性、分類

Ellipse-Space Probabilistic Neural Networks

I-Cheng Yeh
Department of Information Management, Chung Hua University

Kuan-Cheng Lin
Department of Information Management, Chung Hua University

Abstract

This study proposed the ellipse-space probabilistic neural network (EPNN), which includes three kinds of network parameters that can be adjusted through training: the variable weight representing the importance of each input variable of each pattern, the core-width-reciprocal representing the effective range of each pattern, and the data weight representing the reliability of each pattern. These network parameters can improve the accuracy of model, and calculate the variable importance index to offer the ability to appraise importance of each input variable. To prove the performance of EPNN, three artificial classification problems as well as seven actual classification problems were employed to test it and compare it with back-propagation network (BPN) and probabilistic neural network (PNN). The results proved that (1) the accuracy of EPNN is slightly lower than BPN, while strongly higher to PNN in the artificial classification problems; the accuracy of EPNN is obviously higher than BPN and PNN in the actual classification problems, and (2) the variable importance index really expressed the importance of each input variable to the output variable, which makes model have the explanation ability.

Key words : artificial neural network, probabilistic neural network, variable importance, classification.

壹、前言

分類問題是指根據已知類別的樣本建構某種分類架構來判別未知類別的樣本之類別。傳統上常用統計學上的方法作為分類的依據，例如邏輯迴歸、判別分析。這些方法頗具成效，然而在面對變數具有非線性、及變數之間具有交互作用的問題時效果常不理想。類神經網路是非常有效的非線性模型建構工具，它使用大量簡單的相連人工神經元來模仿生物神經網路的能力，能夠對於由外界所輸入的資訊進行學習、回想等一系列動作，因此常被用來處理分類問題。

機率神經網路（Probabilistic Neural Networks, PNN）（Specht 1988; Specht 1990; Burrascano 1991; Patra, et al. 2002）是一種常見的類神經網路模式，其理論基礎是貝氏分類（Bayes classifier）和機率密度函數。假設一分類問題具有k個類別$C_1,C_2,\cdots,C_k$，此一分類問題的分類規則是由m維的特徵向量

$$X=(X_1,X_2,...,X_m) \tag{1}$$

所決定，即在此m維樣本空間中，各分類的機率密度函數為特徵向量的函數$f_1(X), f_2(X),\cdots, f_k(X)$，而貝氏分類器的決策公式：

$$h_i c_i f_i(X) > h_j c_j f_j(X) \text{ 對所有的 } j \neq i \tag{2}$$

其中f_k為第k類的機率密度函數；C_k則代表應為第k類，但被誤判的成本；h_k是第k類的事前機率（prior probability）。

上述機率密度函數一般採用常態機率密度函數（Parzen 1962）：

$$f_a(X)=\frac{1}{(2\pi)^{\frac{2}{m}}\sigma^m}(\frac{1}{n_a})\sum_{p=1}^{n_a}\exp(-\frac{(X-X_{ap})'(X-X_{ap})}{2\sigma^2}) \tag{3}$$

其中$f_a(X)$=當分類A在X點位置的機率密度函數值；p=訓練向量的編號；m=輸入變數的數目；σ=平滑參數；n_a=在A分類中的訓練向量個數；X=測試分類向量；X_{ap}在分類A中第 p 個訓練資料。

因為

$$\frac{1}{(2\pi)^{\frac{2}{m}}\sigma^m}(\frac{1}{n_a})=\text{常數}=h \tag{4}$$

$$(X-X_{ap})'(X-X_{ap})=\sum_{i=1}^{m}(x_i-x_i^{ap})^2 \tag{5}$$

故機率密度函數可簡化為

$$f_a(X) = h\sum_{p=1}^{n_a} f_{ap} \tag{6}$$

$$f_{ap} = \exp(-\frac{\sum_{i=1}^{m}(x_i - x_i^{ap})^2}{2\sigma^2}) \tag{7}$$

其中x_i=樣本的第i個輸入變數的值；x_i^{ap} 樣本庫的A分類的第p個樣本的第i個輸入變數的值。

雖然機率神經網路可以處理分類問題，但無法處理函數映射問題，因此有學者（Specht 1991）以加權平均法將網路的輸出改為

$$y = \frac{\sum_{p=1}^{n} f_p \cdot t_p}{\sum_{p=1}^{n} f_p} \tag{8}$$

其中t_p樣本庫第p個樣本的輸出變數已知值；f_p樣本庫第p個樣本權值；n=樣本庫的樣本數目。

機率神經網路只有一個平滑參數需要調整，不需迭代過程，因此具有快速學習的能力，並且具有一定水準的準確度。但它有二項缺點：（1）在資料量不是很大的情況下，準確度比較差；（2）所建模型缺乏評估輸入變數重要性的能力。

本研究提出橢圓空間機率神經網路（Ellipse-Space Probabilistic Neural Networks, EPNN），它擁有三種可透過訓練來修正的網路參數：代表輸入變數重要性的變數權值、代表樣本有效範圍的核寬倒數、及代表樣本可靠程度的資料權值。其中，變數權值控制機率密度函數的形狀，使得機率密度函數的等高線不再受限為圓形，而可以是橢圓形，而變數權值越大的變數，代表其重要性越大，該方向的橢圓形半徑越小；核寬倒數相當於傳統機率神經網路的平滑參數的倒數，即機率密度函數的寬度之倒數，而核寬越大（核寬倒數越小），代表樣本有效範圍越大；資料權值相當於機率密度函數的高度，高度愈大，代表樣本的可信度越大。這些網路參數可以提升模型的準確度，並提供評估輸入變數重要性的能力。

本研究與其它文獻不同之處與創新的貢獻簡述如下：

一、超橢圓的機率密度函數

傳統的PNN視所有自變數有同等地位，機率密度函數的等高線是圓形。本研究則在機率密度函數中加入「變數權值」，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。此變數權值可視為機率密度函數的「核形狀參數」。本研究也以十個已發表的文獻的例題證明加入形狀參數的機率密度函數確實能大幅提高分類模型的準確度，大幅超越傳統的PNN，甚至比倒傳遞網路更加準確。

二、統一的數學理論架構

傳統的PNN經常以試誤法決定機率密度函數的平滑參數（核半徑參數）。本研究雖然在可調參數上比傳統的RBFN還多了變數權值（核形狀參數）與資料權值（核高度參數），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，包括機率密度函數的核半徑、核形狀、核高度參數。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核形狀參數確實能依題目調整到適當的值。

三、可估計輸入變數重要性的重要性指標

傳統的PNN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「重要性指標」，並透過特殊設計的數值例題證明「重要性指標」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。此外，此指標也可以用來篩選輸入變數，以建構精簡但準確的分類模型。

本文第二節將推導其網路學習規則，第三、四節分別以三個人為的分類問題及七個實際的分類問題來驗證此網路，並與倒傳遞網路及機率神經網路做比較，第五節對整體研究的測試結果做一總結。

貳、理論

一、網路參數的學習演算法

本研究將機率神經網路作如下修正：

$$f_p = h_p^2 \exp\left(-\frac{\sum_{i=1}^{m} W_{ip}^2 (x_i - x_i^p)^2}{2\sigma_p^2}\right) \qquad (9)$$

其中

x_i=未知樣本的第i個輸入變數的值。

x_i^p =樣本庫第p個樣本的第i個輸入變數的值。

W_{ip}=樣本庫第p個樣本的第i個輸入變數權值，它控制機率密度函數的形狀，使得機率密度函數的等高線不再受限為圓形，而可以是橢圓形，而變數權值越大的變數，代表其重要性越大，該方向的橢圓形半徑越小。

h_p=樣本庫第p個樣本的資料權值，它相當於機率密度函數的高度，高度愈大，代表樣本的可信度越大。

σ_p=樣本庫第p個樣本的平滑參數，它相當於機率密度函數的寬度，寬度愈大，代表樣本的有效範圍越大。

為了避免在自適應調整過成中平滑參數可能逼近0，造成上式出現分母為0的問題，故採用核寬倒數代替平滑參數：

令

$$V_p = \frac{1}{\sqrt{2}\sigma_p} \tag{10}$$

其中V_p=樣本庫第p個樣本的核寬倒數，它相當於機率密度函數的寬度之倒數，其值愈大，代表樣本的有效範圍越小。

故

$$f_p = h_p^2 \exp(-V_p^2 \sum_{i=1}^{m} W_{ip}^2 (x_i - x_i^p)^2) \tag{11}$$

令

$$D_p \equiv V_p^2 \sum_{i=1}^{m} W_{ip}^2 (x_i - x_i^p)^2 \tag{12}$$

則

$$f_p = h_p^2 \exp(-D_p) \tag{13}$$

為了導出自適應調整上述參數的公式，令誤差函數

$$E = \frac{1}{2} \sum_{j}^{n} (t_j - y_j)^2 \tag{14}$$

其中n=輸出變數的數目；t_j=訓練範例的第j個輸出變數的已知值；y_j=訓練範例的第j個輸出變數的推論值：

$$y_j = \frac{\sum_p f_p \cdot t_{pj}}{\sum_p f_p} \tag{15}$$

其中y_j=樣本的第j個輸出變數推論值；t_{pj}=樣本庫第p個樣本的第j個輸出變數已知值。

調整參數的目標在於最小化誤差函數，故可依最陡坡降法推導各參數的修正量如下：

（一）變數權值

$$\Delta W_{ip} = -\eta \frac{\partial E}{\partial W_{ip}} \tag{16}$$

其中η=學習速率。

由（14）式以偏微分的連鎖律得

$$\frac{\partial E}{\partial W_{ip}} = \sum_j \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial W_{ip}} \tag{17}$$

由（12）（13）（15）式以偏微分的連鎖律得

$$\frac{\partial y_j}{\partial W_{ip}} = \frac{\partial y_j}{\partial f_p} \frac{\partial f_p}{\partial D_p} \frac{\partial D_p}{\partial W_{ip}} \tag{18}$$

將（18）式代入（17）式得

$$\frac{\partial E}{\partial W_{ip}} = \sum_j \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial f_p}\frac{\partial f_p}{\partial D_p}\frac{\partial D_p}{\partial W_{ip}} \tag{19}$$

其中

由（14）式微分得

$$\frac{\partial E}{\partial y_j} = -(t_j - y_j) \tag{20}$$

由（15）式微分得

$$\frac{\partial y_j}{\partial f_p} = \frac{(\sum f_p)\cdot t_{pj} - (\sum f_p t_{pj})\cdot 1}{(\sum f_p)^2} = \frac{t_{pj} - (\sum f_p t_{pj} / \sum f_p)}{\sum f_p} = \frac{t_{pj} - y_j}{\sum f_p} \tag{21}$$

由（13）式微分得

$$\frac{\partial f_p}{\partial D_p} = -h_p^2 \exp(-D_p) \tag{22}$$

由（12）式微分得

$$\frac{\partial D_p}{\partial W_{ip}} = V_p^2 (2W_{ip}(x_i - x_i^p)^2) \tag{23}$$

將（20）（21）（22）（23）式帶入（19），再帶入（16）式得

$$\Delta W_{ip} = -\eta \frac{\partial E}{\partial W_{ip}} = -2\eta \sum_j (t_j - y_j) \frac{t_{pj} - y_j}{\sum f_p} h_p^2 \exp(-D_p)\cdot V_p^2 \cdot W_{ip} \cdot (x_i - x_i^p)^2 \tag{24}$$

令δ_p為第p個預測輸出值與實際輸出值的差距量

$$\delta_p \equiv \sum_{j=1}^{n} (t_j - y_j)(t_{pj} - y_j) h_p \exp(-D_p) \tag{25}$$

其中t_j=此訓練樣本的第j個輸出變數已知值。

將（24）式簡化為

$$\Delta W_{ip} = -2\eta \frac{1}{\sum f_p} \delta_p h_p V_p^2 W_{ip} (x_i - x_i^p)^2 \tag{26}$$

事實上，（25）式與（26）式是可以用直觀的方式來理解。（25）是有四項構成：（$t_j - y_j$）、（$t_{pj} - y_j$）、h_p、及exp$(-D_p)$，其中第三項與第四項永遠大於0，而第一項與第二項相乘可能大於、等於、或小於0。當相乘大於0時代表二種情形：

情況一：（$t_j - y_j$）>0 且（$t_{pj} - y_j$）>0

前者代表此訓練樣本的第j個輸出變數被低估了，後者代表樣本庫第p個樣本的第j個輸出變數已知值t_{pj}大於此訓練樣本的第j個輸出變數估計值y_j。因此如果提高第p個樣本的權值 W_{ip}，有助於使估計值y_j變大，改善其誤差。

情況二：（$t_j - y_j$）<0 且（$t_{pj} - y_j$）<0

前者代表此訓練樣本的第j個輸出變數被高估了，後者代表樣本庫第p個樣本的第j個輸出變數已知值t_{pj}小於此訓練樣本的第j個輸出變數估計值y_j。因此如果提高第p個樣本的權值W_{ip}，有助於使估計值y_j變小，改善其誤差。

綜合上述二種情況，可知只要（$t_j - y_j$）（$t_{pj} - y_j$）的乘積大於0，無論那一種情況都應該提高第p個樣本的權值來改善其誤差。同理，只要（$t_j - y_j$）（$t_{pj} - y_j$）的乘積小於0，都應該降低第p個樣本的權值來改善其誤差。而公式（25）式顯示，（$t_j - y_j$）（$t_{pj} - y_j$）乘積與差距量δ_p成正比，而公式（26）式顯示，差距量δ_p與第p個樣本的權值W_{ip}成正比，因此可以推得δ_p與權值W_{ip}成正比，正好可滿足上述需求。

（二）資料權值

$$\Delta h_p = -\eta \frac{\partial E}{\partial h_p} \tag{27}$$

由偏微分的連鎖律得

$$\frac{\partial E}{\partial h_p} = \sum_j \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial f_p} \frac{\partial f_p}{\partial h_p} \tag{28}$$

由（13）式微分得

$$\frac{\partial f_p}{\partial h_p} = 2h_p \cdot \exp(-D_p) \tag{29}$$

將（20）（21）（29）式帶入（28），再帶入（27）式得

$$\Delta h_p = -\eta \frac{\partial E}{\partial h_p} = 2\eta \sum_j (t_j - y_j) \frac{t_{pj} - y_j}{\sum f_p} h_p \exp(-D_p) \tag{30}$$

由（25）可將上式簡化為

$$\Delta h_p = 2\eta \frac{1}{\sum f_p} \delta_p \tag{31}$$

（三）核寬倒數

$$\Delta V_p = -\eta \frac{\partial E}{\partial V_p} \tag{32}$$

由偏微分的連鎖律得

$$\frac{\partial E}{\partial V_p} = \sum_j \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial f_p} \frac{\partial f_p}{\partial D_p} \frac{\partial D_p}{\partial V_p} \tag{33}$$

由（12）式微分得

$$\frac{\partial D_p}{\partial V_p} = 2V_p \sum_i W_{ip}^2 (x_i - x_i^p)^2 \tag{34}$$

將（20）（21）（22）（34）式帶入（33），再帶入（32）式得

$$\Delta V_p = -\eta \frac{\partial E}{\partial V_p} = -2\eta \sum_j (t_j - y_j) \frac{t_{pj} - y_j}{\sum f_p} h_p^2 \exp(-D_p) \cdot V_p \cdot \sum_i W_{ip}^2 (x_i - x_i^p)^2 \quad (35)$$

由（25）可將上式簡化為

$$\Delta V_p = -2\eta \frac{1}{\sum f_p} \delta_p \cdot h_p \cdot V_p \cdot \sum_i W_{ip}^2 (x_i - x_i^p)^2 \quad (36)$$

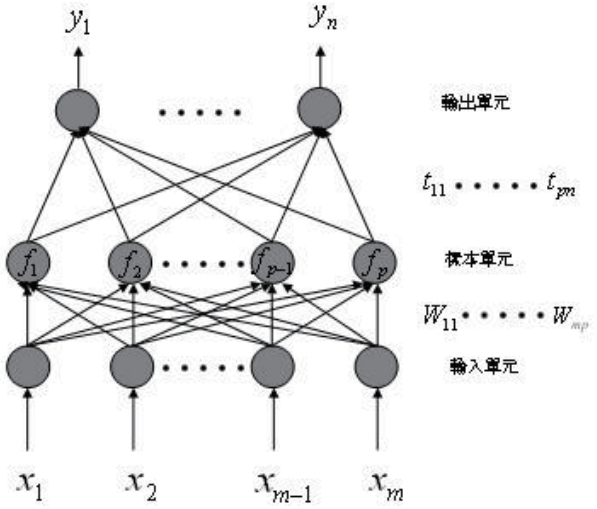

圖1：橢圓空間機率神經網路架構

上述演算法可以使機率密度函數的核之形狀、高度、以及半徑等三種參數在學習過程中被優化，讓每筆樣本的每個輸入變數具有不同的變數權值，使得每筆樣本的機率密度函數都可依訓練樣本而調整成合適的橢圓形，因而減小模型的誤差，因此稱之為「橢圓空間機率神經網路」，其架構如圖1所示。

二、變數重要性的估計

在EPNN中，令第i個輸入變數對第j個輸出變數的一次微分如下：

$$\frac{\partial y_j}{\partial x_i} = \sum_p \frac{\partial y_j}{\partial f_p} \frac{\partial f_p}{\partial D_p} \frac{\partial D_p}{\partial x_i} \quad (37)$$

由（12）式微分得

$$\frac{\partial y_j}{\partial x_i} = \sum_p \frac{\partial y_j}{\partial f_p} \frac{\partial f_p}{\partial D_p} \frac{\partial D_p}{\partial x_i} \quad (38)$$

將（21）（22）（38）式帶入（37）式得

$$\frac{\partial y_j}{\partial x_i} = \sum_p \frac{y_j - t_{pj}}{\sum f_p} h_p^2 \exp(-D_p) \cdot V_p^2 \cdot W_{ip}^2 \cdot (x_i - x_i^p) \quad (39)$$

因為只要輸入變數對輸出變數具有影響力，則其一次微分值必定顯著異於0，但由公式（39）可知，一次微分值與樣本的輸入變數值有關。因此為了消除個別樣本的差異，可將所有訓練範例用（39）式算出其一次微分值，再以下式估算變數的重要性：

$$I_i^j = \sqrt{\frac{\sum_{k=1}^{P}\left(\frac{\partial y_j}{\partial x_i}\Big|_k\right)^2}{P}} \qquad (40)$$

其中 I_i^j 第j個輸出變數的第i個輸入變數的「重要性指標」；P＝訓練範例數目。

雖然上式只含一次微分，但它將一次微分取平方和，因此無論自變數對因變數是正相關、負相關、曲線相關（開口朝上或朝下）、或與其它變數組成交互作用，其一次微分的平方和均會顯示出大的數值，表示此自變數對因變數具有重要性。只有當自變數對因變數是不相關時，其一次微分的平方和才會顯示出很小的數值，表示此自變數對因變數不具有重要性。當「重要性指標」越大，代表輸入變數的作用越顯著。因為這一個指標可以定量衡量每一個輸入變數的重要程度，因此在網路訓練完畢後，計算此指標可以使EPNN在具有預測能力之外，也能具有判斷變數重要性的能力，提高其模型的透明度，改善其黑箱模型的缺點。

三、EPNN學理上的潛在優點

（一）兼具PNN與BPN的優點

EPNN可視為以PNN為架構，以與BPN相同的最小化誤差函數方法優化神經網路，故兼具二者的優點。PNN的優點是因為它是一種lazy learning演算法，只有平滑參數（核半徑參數）需要調整。因為只有單一參數，因此要決定其最佳值十分快速。BPN的優點是因為它是一種eager learning演算法，使用最小化誤差函數方法優化網路，雖然優化過程耗時，但經常比PNN準確。而EPNN雖然有三種參數需要調整，但在開始學習前可先令變數權值（核形狀參數）與資料權值（核高度參數）均為1.0，並令各樣本的核寬倒數使用同一個值，如此只要調整核寬倒數這一個參數，EPNN就具有與PNN同級的預測能力。再以與BPN相同的最小化誤差函數方法調整這三種參數，EPNN可以達到接近BPN同級的預測能力。

（二）具有漸增學習的優點

BPN並未保存各樣本的特定資訊在網路的架構中，因此當有新樣本輸入系統時，必須重新學習全體樣本，對於樣本不斷更新的問題將耗費大量的運算時間。而EPNN具有lazy learning演算法的架構，因此各樣本的特定參數（變數權值、資料權值、核寬倒數）仍保留在網路的架構中。當有N個新樣本輸入系統時，可以排除n個最舊的樣本的特定參數，然後在其餘舊樣本的特定參數不變下，只調整此n個新樣本的特定參數。因此EPNN可以進行漸增式學習（incremental learning）。

參、數值例題

本研究設計了三個不同類型的函數來測試與探討EPNN的性能與特性。這三種例題分別有十個輸入變數X_1～X_{10}，及二個代表二個不同類別的二元輸出變數Y_1、Y_2。每一個例題的輸入變數皆在0～1的範圍內隨機取點，產生1000個訓練範例與100個測試範例。這三組資料將用來比較倒傳遞網路（BPN）、機率神經網路（PNN）與橢圓空間機率神經網路（EPNN）的準確度。此三個例題也將用來驗證重要性指標是否可以顯示輸入變數影響輸出變數的重要程度。

為得到最佳結果，在此採用不同的網路架構、學習參數，以建立最適模型：

- BPN：嘗試不同的隱藏層單元數目（4, 8, 16, 32）與學習速率（0.1, 0.3, 1.0, 3.0）。
- PNN：嘗試不同的平滑參數（0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0）。
- EPNN：嘗試不同的學習速率（0.1, 0.3, 1.0, 3.0）與初始核寬倒數（0.1, 0.3, 1.0, 3.0）。

一、例題一：線性與二次函數（分類）

假設有如下分類函數

$$Y = 0X_1 + 1X_2 - 2X_3 + 4X_4 - 8X_5 + 4[0(X_6 - 0.5)^2 - 1(X_7 - 0.5)^2 + 2(X_8 - 0.5)^2 - 4(X_9 - 0.5)^2 + 8(X_{10} - 0.5)^2] \quad (41)$$

如果$Y>0$則為第一類；否則為第二類。

在上式中，可明顯看出X_1～X_5和輸出變數的關係為線性且變數重要性逐漸上升，X_6～X_{10}和輸出變數的關係為二次且變數重要性逐漸上升。結果顯示，在不同的網路架構、學習參數下，BPN、PNN、EPNN所建立的最適模型的測試範例誤判率分別為0.05、0.19、0.08，顯示EPNN的準確度只略低於BPN，而遠優於PNN。

圖2顯示例題一的重要性指標結果，可見 X_1～X_5及X_6～X_{10}對輸出變數的重要性逐漸上升，其中，X_5、X_{10}的指標值相對較高，代表此變數極為重要，與此人為函數的應有結果一致，證實了的重要性指標確實可以衡量線性及二次分類函數的輸入變數之重要程度。

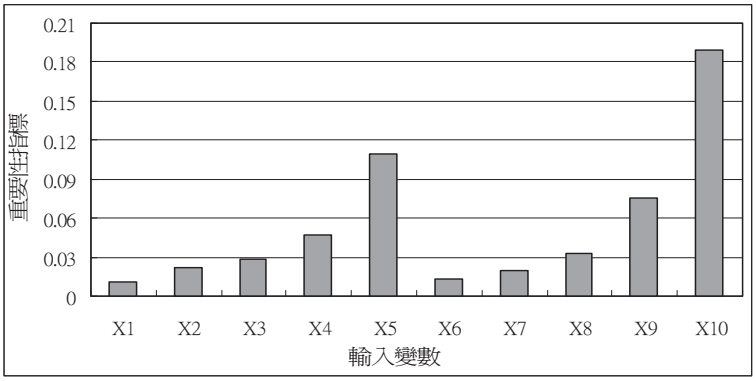

圖2：例題一的重要性指標

二、例題二：交互與二次函數（分類）

假設有如下函數

$$Y = 1.5(X_1 - 0.5)(X_8 - 0.5) + 1.5(X_4 - 0.5)(X_{10} - 0.5) + (X_5 - 0.5)^2 \quad (42)$$

如果$Y>0$則為第一類；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。結果顯示，在不同的網路架構、學習參數下，BPN、PNN、EPNN所建立的最適模型的測試範例誤判率分別為0.05、0.26、0.07，顯示EPNN的準確度與BPN相近，而遠優於PNN。

圖3顯示例題二的重要性指標結果。由圖可知，X_1、X_4、X_5、X_8、X_{10}具有較大的指標值，代表它們是較重要的變數，而X_2、X_3、X_6、X_7、X_9具有較小的指標值，代表它們是較不重要的變數。此結論與前述例題二的應有結果一致，證實了重要性指標可以衡量混合交互作用、曲率作用的分類函數之輸入變數重要程度。

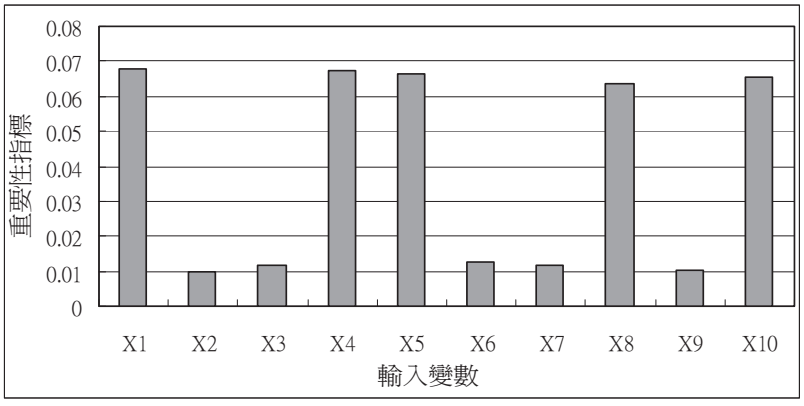

圖3：例題二的重要性指標

三、例題三：混合函數（分類）

假設有如下函數

$$Y = (X_1 - 0.25)^2 - (X_3 - 0.75)^2 + X_5 - X_7 + (X_9 - 0.25) * (X_{10} - 0.75) \quad (43)$$

如果$Y>0$則為第一類；否則為第二類。

上式函數中，X_1、X_3為曲率作用變數，X_5、X_7為線性作用變數，X_9、X_{10}為一組交互作用變數，X_2、X_4、X_6、X_8為無關變數。結果顯示，在不同的網路架構、學習參數下，BPN、PNN、EPNN所建立的最適模型的測試範例誤判率分別為0.01、0.08、0.06，顯示APNN的準確度低於BPN，而優於PNN。

圖4顯示例題三的重要性指標結果。由圖可知，X_1、X_3、X_5、X_7、X_9、X_{10}具有較大的指標值，代表它們是較重要的變數，而X_2、X_4、X_6、X_8為具有較小的指標值，代表它們是較不重要的變數。此結論與前述例題三的應有結果一致，證實了重要性指標確實可

以提升衡量混合線性作用、交互作用、曲率作用的分類函數的輸入變數重要程度。

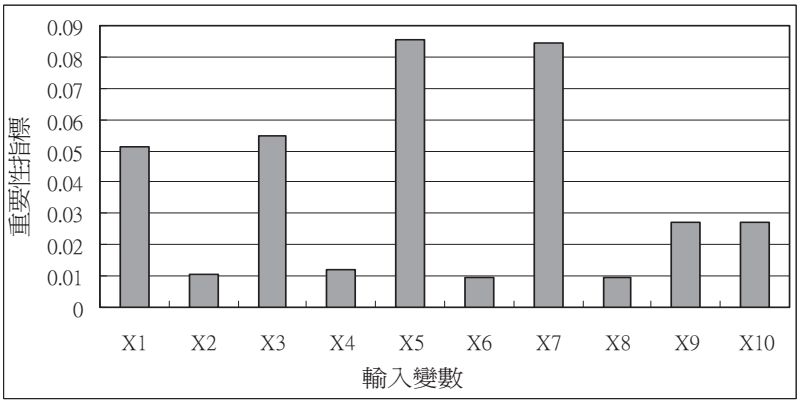

圖4：例題三的重要性指標

四、誤判率之比較

將三個數值例題的誤判率，繪圖比較如圖5。可見在例題一與例題二中EPNN的誤判率明顯接近BPN，而遠低於PNN；在例題三中EPNN的誤判率雖明顯高於BPN，但仍遠低於PNN，證實了EPNN的準確性遠優於PNN。這可能是因為在學習過程中，EPNN不像PNN採取固定的變數權值，而可以調整各樣本的變數權值、核寬倒數、資料權值，使其機率密度函數的形狀能夠更接近分類問題的本質，而使模型有更小的誤差。

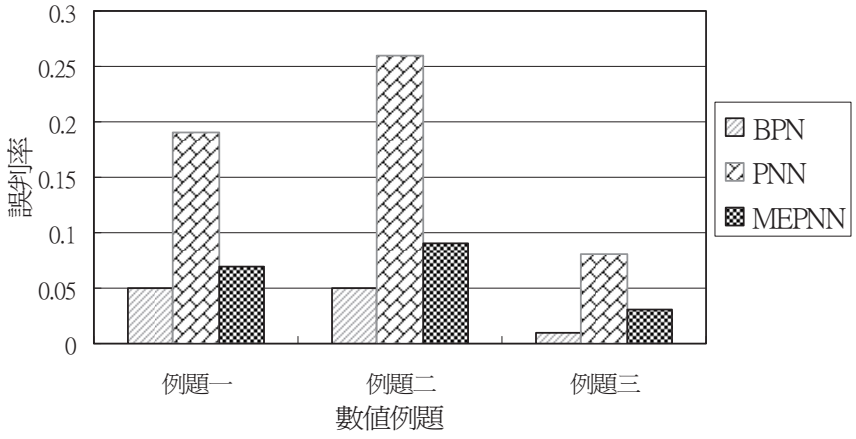

圖5：三個數值例題的誤判率之比較

五、EPNN可以依照樣本點調整變數權值嗎？

在EPNN中，每個樣本都有特定的變數權值，在輸入變數很多，且樣本很多的情況下，會有大量的變數權值。為了證明EPNN可以依照樣本點所處位置的特性，調整變數

權值，在此假設有如下函數

$$Y = X_1 \cdot X_2 \tag{44}$$

此函數的等高線圖如圖6。由圖可知，左上方處的Y值與X_1的大小較相關，與X_2的大小較不相關，因此在計算樣本間的「距離」時，X_1的權重應大於X_2的權重，故W_{1P}/W_{2P}會較大。同理，在右下方處的Y值與X_1的大小較不相關，與X_2的大小較相關，因此在計算樣本間的「距離」時，X_1的權重應小於X_2的權重，故W_{1P}/W_{2P}會較小。在左下到右上的對角線上，Y值與X_1及X_2的大小之相關性相同，因此X_1的權重應等於X_2的權重，故W_{1P}/W_{2P} =1.0。

假設各輸入變數皆在0～1的範圍內，在此範圍內隨機取樣，產生300個訓練範例與100個測試範例。以EPNN建模後，將第p個訓練範例樣本的X_1與X_2的變數權值的相對大小W_{1P}/W_{2P}與W_{1P}/W_{2P}分別繪成圖7、圖8的泡泡圖，其中變數權值的比值與泡泡的寬度成比例。由圖7可知，W_{1P}/W_{2P}在圖的左上方最大，右下方最小。由圖8可知，W_{2P}/W_{1P}在圖的左上方最小，右下方最大。在圖7、圖8的左下到右上的對角線上，W_{1P}/W_{2P}、W_{2P}/W_{1P}近似1.0。此一現象與上述圖6的說明相符，由此可見EPNN演算法確實可以依照樣本點所處位置的特性，調整變數權值。

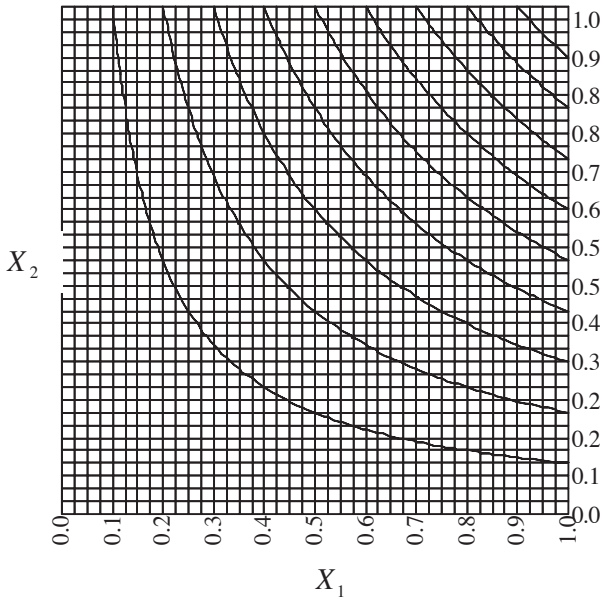

圖6： $Y = X_1 \cdot X_2$ 函數的等高線圖

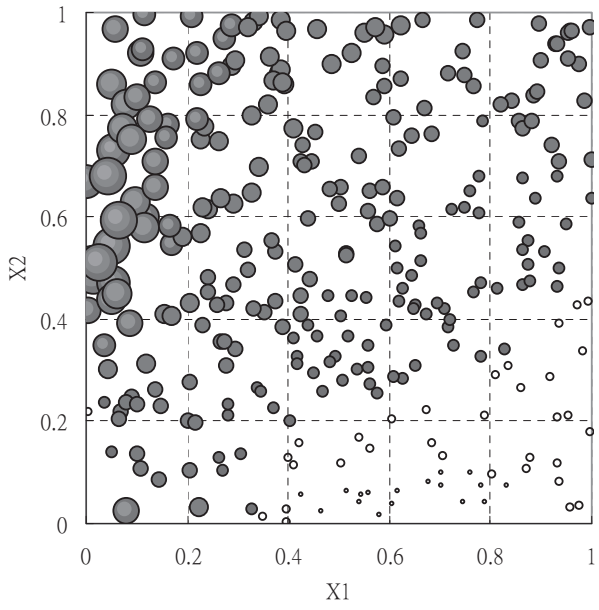

圖7：W_{1P}/W_{2P}的泡泡圖

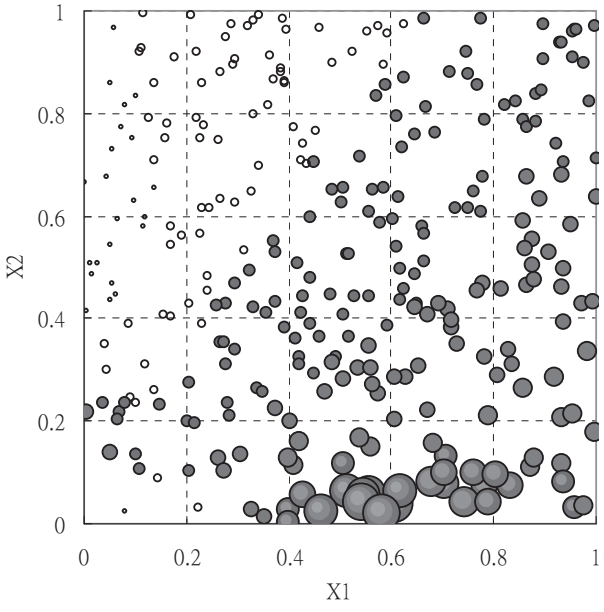

圖8：W_{2P}/W_{1P}的泡泡圖

肆、應用實例

一、七個實際分類問題之測試與比較

為了解EPNN在分類問題上的效能，本研究將以健身中心會員開發（文少宣2004）、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）、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）、遙測影像分類（UCI Machine Learning Repository Content Summary 2007）及垃圾郵件過濾（UCI Machine Learning Repository Content Summary 2006）等七個實際分類例題來測試，並與BPN、PNN的準確度做一比較。

在此，同前節一樣，採用不同的網路架構、學習參數，以建立最適模型。誤判率的數據結果顯示如表1。由表可知，EPNN的模型最準確，其次是BPN，最差的是PNN。

為了更清楚地比較各網路架構在每一例題上的優劣，將誤判率結果整理如圖9。由圖可知，EPNN除了例題一（健身中心會員開發）的誤判率略高於BPN，其餘實際分類例題皆低於BPN與PNN。

表1：七個實際分類例題的測試範例誤判率

	例題名稱	BPN	PNN	EPNN
1	健身中心會員開發	0.328	0.392	0.371
2	休旅車潛在顧客開發	0.238	0.163	0.138
3	汽車保險潛在顧開發	0.340	0.336	0.314
4	集集大地震引致山崩	0.229	0.213	0.195
5	森林地表覆蓋分類	0.214	0.254	0.194
6	遙測影像分類	0.120	0.100	0.072
7	垃圾郵件分類	0.062	0.123	0.051
	平均	0.219	0.226	0.191

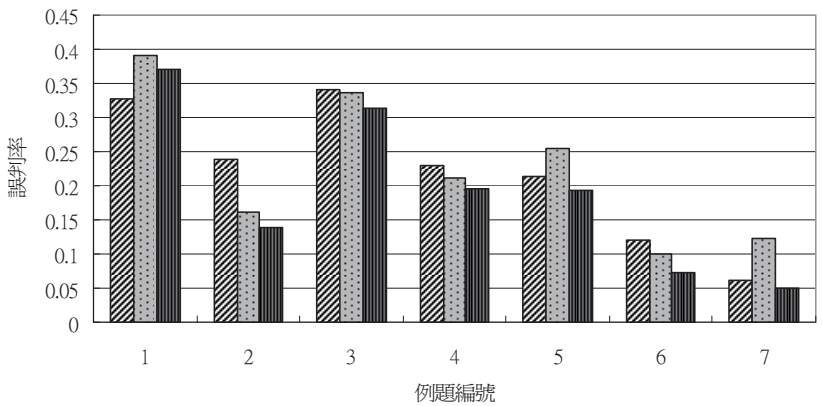

圖9：七個實際例題之誤判率比較

二、變數重要性之探討

為了證明EPNN演算法的重要性指標是否可以提升正確判斷輸入變數對輸出變數影響的重要程度，在此以前述的森林地表覆蓋分類實例來進行測試。在此資料集中，森林被分割 EMBED Equation.3 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2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3。其中3000筆做為訓練範例，1000筆做為測試範例。

表2：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=高程	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3：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

上述資料集以EPNN用最佳學習參數建模後，將各輸入變數對各輸出變數 的重要性指標取平均值，結果如圖10。較重要的變數將會有較大的平均值。由圖可知，前五個較為最重要的輸入變數分別是高程(X_1)、對道路水平距離(X_6)、對火點距離(X_{10})、對水體水平距離(X_4)、對水體垂直距離(X_5)，其重要性明顯高於其他輸入變數。

為了證明這五個輸入變數確實是影響分類的最重要變數，在此選取此五個變數建立EPNN分類模型，再以隨機方式從全部14個變數中抽出30組五個變數，分別建立模型。結果如圖11所示，取前五個最重要的輸入變數所建立的分類模型其測試範例誤判率為0.194，而取隨機選出的五個輸入變數者其誤判率最小值為0.211，最大值為0.61，平均值為0.394，標準差為0.102。依單尾t統計檢定，依重要性指標選取五個變數所建立的模型其誤判率低於依隨機選取者的顯著性為0.03，證實重要性指標確實可以找出影響輸出變數的最重要輸入變數。

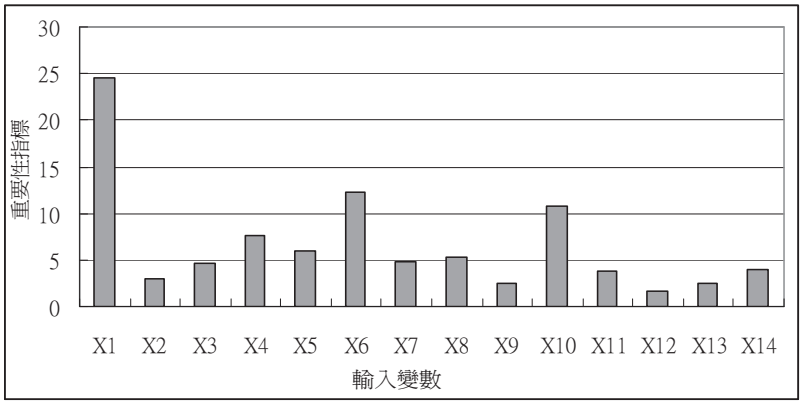

圖10：森林地表覆蓋類型實例的重要性指標

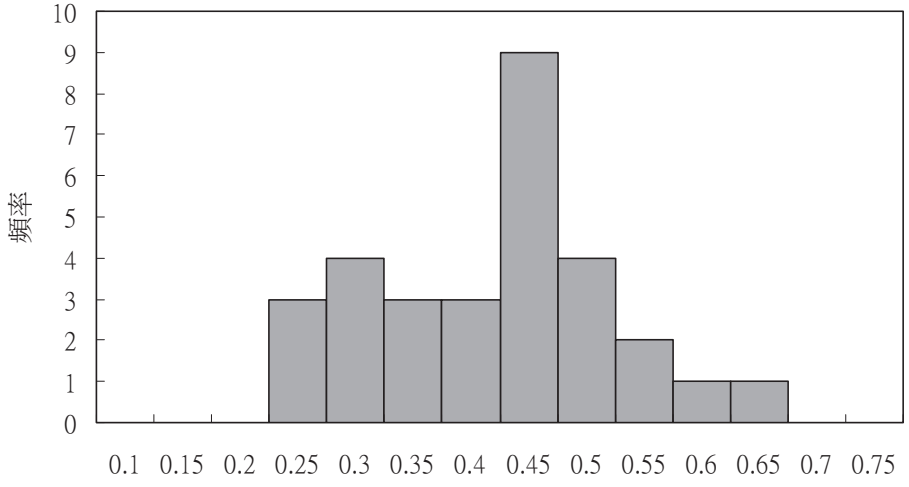

圖11：以五個輸入變數所建立的EPNN分類模型的測試範例誤判率直方圖

伍、結論

本研究提出橢圓空間機率神經網路，它擁有三種可透過訓練來修正的網路參數：代表輸入變數重要性的變數權值、代表樣本有效範圍的核寬倒數、及代表樣本可靠程度的資料權值。本研究以三個人為的分類問題以及七個實際的分類問題來做測試，並與倒傳遞網路（BPN）及機率神經網路（PNN）做比較。得到下列結論：

一、在人為的分類問題EPNN的模型準確度只略低於BPN，而遠高於PNN；在實際的分類問題EPNN的模型準確度明顯高於BPN與PNN。可見EPNN可以改善PNN的準確度。

二、EPNN的重要性指標可以合理地估計各輸入變數影響輸出變數的重要程度，使模型具有解釋能力。

誌謝

本研究承蒙國科會贊助（計畫編號97-2221-E-216-038），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班論文。
2. 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。
3. 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.
4. Burrascano, P.“Learning vector quantization for the probabilistic neural network,”IEEE Transaction on Neural Networks(2:4), 1991, pp. 458-461.
5. Megaputer Intelligence, Inc. PolyAnalyst Case Studies, 2007, (http://www.megaputer.com/)
6. Parzen, E.“On estimation of a probability density function and mode,”Annals of Mathematical Statistics(33:3), 1962, pp. 1065-1076.
7. Patra, P. K., Nayak, M., Nayak, S. K., and Gobbak, N. K.,“Probabilistic Neural Network for Pattern Classification,”Proceedings of the 2002 International Joint Conference on Neural Networks, Honolulu, 2002, Vol. 2, pp. 1200-1205.
8. Specht, D. F.“Probabilistic neural networks for classification, or associative memory,”Proceedings of the 1988 IEEE International Conference on Neural Networks, San

Diego,1988, Vol. 1, pp. 525-535
9. Specht, D. F.“Probabilistic Neural Networks,”Neural Networks(3:1), 1990, pp. 109-118.
10. Specht, D. F.“A General Regression Neural Network,”IEEE Transactions on Neural Networks(2:6), 1991, pp. 568-576.
11. The Insurance Company Benchmark(COIL 2000), 2007, http://kdd.ics.uci.edu/databases/tic/tic.html
12. UCI Machine Learning Repository Content Summary, Spambase Database, 2006, (http://www.ics.uci.edu/~mlearn/MLSummary.html)
13. UCI Machine Learning Repository Content Summary, Statlog Project Databases, 2007, (http://www.ics.uci.edu/~mlearn/MLSummary.html)