

# 應用類神經網路構建預測侵台颱風次數之模型研究

殷 富  
文化大學應用數學系 教授

## 摘要

颱風對台灣來說，可能會帶來雨量或強風而使缺水的水庫有雨水氾濫，因而成為台灣水庫重要蓄水來源之一，它使酷熱的夏季，有了幾天涼爽天氣，但也可能造成風災、水災、土石流等而帶來生命或財產的損失與傷亡。颱風成為台灣又愛又恨的大自然力量，對日常生活及經濟活動影響甚鉅，如能對每年侵台颱風次數做準確預測，則能對水庫運轉水資源利用作合理有效的規劃、農業產銷、各種工程執行與施工及規劃等，以減少颱風對台的負面衝擊。

由於太陽與地球運轉的相對位置的關係，使地球接受太陽的日照不均，年平均在地球赤道附近最多，向南北兩極趨減。為使地球各緯度的接受的熱能，能達到平衡，必使接受熱能較多的低緯度熱能帶到高緯度去才可達成，其主要動力有：大氣環流、季風、海洋環流、降水以及颱風等。本研究假設颱風為其餘動力輸送熱能帶到高緯度後仍不足的最後促使地球熱平衡的動力機制，也就是其餘動力輸送熱能帶到高緯度後仍不足的数量較大時，則接下來颱風肩負的使命則大些，也就是颱風發生的次數要多，反之亦然。

呈現輸送熱能帶到高緯度的因子有：海水水溫、氣溫、降水、氣壓、風及聖嬰現象反聖嬰現象等，經以上諸因子與颱風生成總數相關性分析後，本研究採用海水水溫、降水及前一年的產生的颱風能量來預測未來可能發生侵臺颱風數。一般分析方法，不易將此眾多因子同時納入，並消除資料間之干擾雜訊。本研究即運用類神經網路的容錯及消除資料間之干擾能力的特性，來建立預測侵台颱風次數的模型，又因部分因子，其早期資料不易獲得，本研究僅採用 22 年的資料來進行分析。為獲得較佳的結果，本研究將運用可處理僅有少量資料之交叉驗證法(Cross Fold Validation method)來進行分析研究。倒傳遞網路用來建立模型，其輸入變數採用 6 個，輸出變數則為當年全年侵台颱風次數。

本模型之訓練過程及測試過程所使用之輸入及輸出資料是採用 1982 年至 2004 年收集之資料，運用能有效處理少量資料之交叉驗證法，將全部資料分批作為測試驗證之用，經調整模型網路的隱藏層的處理單元數目，學習速率與慣性因子之大小後，得出最佳的模型隱藏層為 1 層有 9 個人工神經元測試運算之誤差均方根平均為 0.1164，侵台颱風之觀測值與計算值之誤差在 1.5 個以內。顯示應用本模型來預測侵台颱風次數是具體可行的。經相關資料輸入本模型後，計算出 2005 年侵臺颱風數為 5.7 個，較 2004 年要之 8 個要少。

關鍵詞：颱風，類神經網路

## 一、前言

侵台颱風可能會帶來雨量或強風使缺水的水庫雨水氾濫，成為台灣水庫重要蓄水來源之一，使酷熱的夏季，有了幾天涼爽天氣，也可能造成風災水災土石流帶來財產的損失，生命的傷亡。颱風成為台灣又愛又恨的大自然力量，對日常生活及經濟活動影響甚鉅，如能對每年侵台颱風數準確預測，則能對水庫運轉，水資源利用，作合理有效的規劃，農業產銷，各種工程規劃與施工及早規劃等，以減少颱風的負面衝擊。

發生在太平洋和南中國海的熱帶氣旋，如中心最大風速在 17.2m/sec 以上，即稱為颱風，依中

心最大風速之大小區分為在 17.2-32.6m/sec，32.7-50.9m/sec 及 51m/sec 以上，分別稱為輕度颱風，中度颱風及強度颱風等三級。根據 Henderson-Sellers et al(1998)分析由 1970 至 1995 年的颱風資料統計，全球每年平均約有總颱風數為 86.1 個，其標準差為 7.9 個，而發生在對台灣可能有影響的西北太平洋有 27 個，其標準差為 3.9 個，也就是西北太平洋的颱風發生數約佔全球總數的三分之一。蕭(1999)認為侵台颱風次數，平均每年約有 4.5 個，最多 9 個，最少 1 個，每年變化率約 30%。

Gray(1997)認為海洋蘊藏的能量是颱風在形成之前，與隨後發展的主要能量來源。崔等(2003)分析聖嬰現象時，發現聖嬰年時，臺灣降水減少，而反聖嬰年時臺灣降水增加，在西北太平洋及南海地區颱風強度減弱。蕭(2000)研究發現在西太平洋，

當聖嬰現象發生時或反聖嬰現象發生時，颱風生成總數，呈現減少的趨勢。

由於太陽與地球運轉的相對位置的關係，使地球接受太陽的日照不均，年平均在地球赤道附近最多，向南北兩極趨減。為使地球各緯度的接受的熱能，能達到平衡，必使接受熱能較多的低緯度熱能帶到高緯度去才可達成，其主要動力有：大氣環流、季風、海洋環流、降水以及颱風等。本研究假設颱風為其餘動力輸送熱能帶到高緯度後仍不足的最後促使地球熱平衡的動力機制，也就是其餘動力輸送熱能帶到高緯度後仍不足數量較大時，則接下來颱風肩負的使命則大些，也就是颱風發生的次數要多，反之亦然。

呈現輸送熱能帶到高緯度的因子有：水溫，氣溫，降水，氣壓，風及聖嬰現象反聖嬰現象等，經諸因子與颱風生成總數相關性分析後，本研究採用水溫降水及前一年的產生的颱風能量來預測未來可能發生侵臺颱風數。一般分析方法，不易將此眾多因子同時納入，並消除資料間之干擾雜訊。本研究即運用類神經網路的容錯及消除資料間之干擾能力的特性，來建立預測侵台颱風次數的模型，又因部分因子，其早期資料不易獲得，本研究僅採用 22 年的資料來進行分析。為獲得較佳的結果，本研究將運用可處理僅有少量資料之交叉驗證法(Cross Fold Validation method)(Su 2005)來進行分析研究。

## 二、類神經網路

### 2.1. 模型架構

類神經網路(Artificial Neural Network)為透過對生物神經模仿學習的一種資訊處理系統，具有將輸入變數與輸出變數間，複雜的內在對映關係充分呈現的功能。其為解決非線性動態問題的最佳工具之一，其由許多人工神經細胞所組成，人工神經細胞亦可稱為處理單元，每一個處理單元的輸出以扇形送出，成為其它處理單元的輸入。

本研究所用之倒傳遞網路(back propagation network, BPN)為目前應用最廣的類神經網路，倒傳遞網路基本原理是利用最陡坡降法(Gradient Steepest Descent Method)的觀念將誤差函數予以最小化，其架構是由數層互相連結的人工神經元所組成：包括輸入層(Input layer)，輸出層(output layer)及若干隱藏層(hidden layer)。各層包含了若干神經元，類神經網路即依學習法則，運用以往實例，透過訓練以調整連結鍊的加權值的方式來完成學習目標的收斂，其架構示於圖 1。

各層之人工神經細胞亦可稱為處理單元，每一個處理單元的輸出，以扇形送出，成為其它處理單元的輸入，處理單元其輸出與輸入的關係式，可用輸入值加權值積和之函數表示，其人工神經元之模型示於圖 2，輸出變數  $y$  與輸入變數  $x$  的

關係式，可用輸入值加權和之函數表示：

$$y = f(w x + b)$$

式中

$y$ : 為輸出變數

$f$ : 為神經元的轉換函數(transfer function) 其為將輸入值之加權乘積和轉換成處理單元輸出值

$w$ : 為連結加權值

$x$ : 輸入變數

$b$ : 為神經元的偏權值(bias)

### 2.2. 模型建立程序

將實例資料分割為學習(或稱訓練)過程資料，及回想(或稱測試)過程資料兩部分：學習過程為運用實例資料，調整網路中之權重值(即  $w$  與  $b$  值)。其流程示於圖 3，為建立適合的網路模型參數，回想過程為將實例資料成為測試案例或待推案例，以案例之輸入變數值，代入學習過程所建立的網路，計算出輸出變數。以測試案例來推估本網路之誤差度，待推案例用以推算其預測值。

為瞭解學習品質，設能量函數或稱誤差函數  $E$  表示如下：

$$E = (1/2) \sum_j (T_j - A_j)^2$$

上式中：

$T_j$ : 輸出層目標輸出值

$A_j$ : 輸入層目標輸入值

因此網路的學習過程即為使上述能量函數最小化的過程，通常以最陡坡降法來使能量函數最小化，及每當輸入一個訓練範例，網路即小幅調整加權指數的大小，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大小成正比。如下列公式所示：

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}}$$

上式中：

$W_{ij}$ : 介於第  $n-1$  層的第  $i$  個處理單元，與第  $n$  層的第  $j$  個處理單元間的連結加權值。

$\eta$ : 學習速率(Learning Rate)，其為控制每次以最陡坡降法最小化誤差函數的步幅。

不論  $W_{ij}$  是否處於輸出層與隱藏層之間， $\frac{\partial E}{\partial W_{ij}}$  之

通常是可寫成下式：

$$\frac{\partial E}{\partial W_{ij}} = -\delta_j^n A_j^{n-1} \quad \text{此式中：}$$

$A_j^{n-1} = W_{ij}$  所連接之較低層處理單元的輸出值

$\delta_j^n = W_{ij}$  所連接之較上層處理單元之差距量

如果  $W_{ij}$  處於輸出層與隱藏層之間,  $W_{ij}$  也就是所連接上層處理單元為輸出層,  $\delta_j^n$  可表示為以下公式:

$$\delta_j^n = (T_j - A_j) f'(net_j)$$

上式中:

$$net_j = \sum_i W_{ij} A_i - b_j$$

若  $W_{ij}$  非處於輸出層與隱藏層之間, 也就是  $\delta_j^n$  所連接層處理元為輸出層表示為:

$$\delta_j^n = (\sum_k \delta_k^{n-1} W_{kj}) f'(net_j^n)$$

若將公式  $\frac{\partial E}{\partial W_{ij}} = -\delta_j^n A_j^{n-1}$  代入

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} \text{ 即得下式:}$$

$$\Delta W_{ij} = \eta \delta_j^n A_i^{n-1}$$

該公式即為倒傳遞網路最陡坡降法演算式之調整權重式。

### 三、模型之建立

本研究為運用倒傳遞網路系統所建立之預測當年發生侵台颱風次數模式, 旨在於利用非線性模式進行氣候, 水溫因子與發生侵台颱風次數間的複相關回歸關係訓練。本研究採用葉怡成(2001)開發之倒傳遞網路軟體進行分析與探討, 據以測式修正網路模式, 進而降低整體模式之錯誤均方差。本研究網路在網路學習循環達 6,000 次, 學習速率為 8.8 時, 整體網路訓練可達最佳收斂效果, 而當網路隱藏層之層數為 1 層及人工神經元數目為 9 時, 其訓練, 測試及錯誤均方差可降至最低(見圖 4)。

由於早期之人造衛星收集水溫資料取得不易, 本研究僅以 22 組資料(1983 至 2004 年)供模型測試之用, 因此本研究之倒傳遞網路測試方式為採用可供少量資料分析的摺疊驗證(Fold Validation method)中之交叉驗證法(Cross Validation method)(Su 2005), 即為將收集的 22 組資料分為 4 次網路模式測試, 其中 3 次為每次分別取 6 組資料作測試, 其餘 1 次為取 4 組資料作測試, 將 22 組資料分 4 次輪流測試完畢, 而在每次進行網路模式測試時, 所運用的學習(訓練)資料即為 22 組資料中未被用為測試之剩餘資料。本模型之輸入變數, 輸出變數及網路參數等三部分分述如下:

#### 1. 輸入變數

依可能影響侵台颱風數的可收集的因子, 定出 12 個輸入變數, 分述如下:

- (1). 當年 1 月至 4 月之台灣南北雨量:  
台灣南北雨量指基隆台北高雄與恆春等地之中央氣象局氣象觀測站之雨量觀測值。
- (2). 當年 1 月至 4 月之台灣北南雨量差:  
台灣北南雨量差為指基隆與台北中央氣象局氣象觀測站之雨量觀測值之和, 減去高雄與恆春中央氣象局氣象觀測站之雨量觀測值之和。
- (3). 前一年全年全臺灣之雨量和:  
全臺灣指基隆, 台北, 台中, 恆春與花蓮等之中央氣象局氣象觀測站之雨量觀測值。
- (4). 前一年北太平洋東西水域海表面水溫差:  
北太平洋東西水域指 120° W, 24° N 與 122° E, 24° N 二處。
- (5). 當年一月至四月北太平洋東西水域海表面水溫差:  
北太平洋東西水域指 120° W, 24° N 與 122° E, 24° N 二處。
- (6). 前一年侵台颱風之颱風總能量參數:  
颱風能量參數指將輕度颱風發生次數乘以 1 倍, 加上中度颱風發生次數乘以 2.5 倍, 再加上強烈颱風發生次數乘以 4 倍後之計算值。  
將第(1)至第(6)個輸入變數數值分別取對數後的數值, 即為第(7)至第(12)個輸入變數。  
以上之雨量及侵台颱風數為取自中央氣象局氣公告之資料, 北太平洋東西水域海水表面水溫資料, 為取自美國國家環境預測中心(NCEP)之人造衛星觀測資料。

#### 2. 目標變數(target variable)

本研究之目標變數為當年度侵台颱風次數, 該資料取自中央氣象局氣公告之每年侵台颱風次數資料。

#### 3. 網路參數

本研究因受制於北太平洋東西水域之海表面水溫資料之取得受限, 故本研究採用之資料為 1982 年至 2004 年, 共計 22 年。其網路參數之建立, 為依圖 3 之學習過程之流程執行, 本研究模型建模之網路參數說明如下:

(1) 學習(訓練)過程資料: 共分 4 次, 其中 3 次每次用 16 組資料, 其餘 1 次用 18 組資料代入計算。

(2) 回想(測試)過程資料: 共分 4 次, 其中 3 次每次用 6 組資料, 其餘 1 次用 4 組資料代入計算。

(3) 網路參數: 用本學習過程資料, 依圖 3 之網路學習過程之流程圖中之流程, 調整其網路權重值及隱藏層之層數及人工神經元數目, 以使案例目標值與案例經模擬計算後之輸出值, 其錯誤均方根(Error of Root Mean Square)為最小, 本研究之最佳網路參數如下:

輸入層: 12 個人工神經元

輸出層:1 個人工神經元

隱藏層:共 1 層有 9 個人工神經元,轉換函數採用雙彎曲函數(Sigmoid function)函數即

$$f(x)=1/(1+\exp(-x))$$

### 3.4. 資料尺度化處理

為配合傳輸函數之範圍在[-1,1]之間,須對資料變數予以尺度化處理,其方法說明如下:

(a) 輸入變數尺度化處理:

設變數之平均值為 M,標準差為 S,原資料值為 Vd,尺度化處理後之資料值為 Vn,其 Vn 計算公式如下:

$$V_n = (V_d - M) / (1.96 S)$$

經過上式之尺度化處理後,約有 95%的資料變數可映射到[-1,1]區間。

(b) 輸出變數尺度化處理

設變數之最大值為 Dx,最小值為 Dn,原資料值為 Vd,尺度化處理後之資料值為 Vn,其 Vn 計算公式如下:

$$V_n = 0.6(V_d - D_n) / (D_x - D_n) + 0.2$$

經過上式之尺度化處理後,資料變數可映射到[0.2, 0.8]區間。

## 四. 討 論

### 4.1. 模型之測試

將以上模型架構,使用交叉驗證法,將測試資料分為 4 組即:第一組資料為 1983 至 1988 年,第二組資料為 1989 至 1994 年,第三組資料為 1995 至 1998 年及第四組資料為 1999 至 2004 年,22 組資料除去測試資料後其餘即為訓練資料,將訓練資料及測試資料代入模式運算,得出其測試之誤差均方根分別為 0.1234, 0.1101, 0.0969 及 0.135,而 1983 至 2004 年之實際每年侵台颱風次數與本模型計算值,示於圖 5。由此圖可知其模型模式運算誤差均控制在 1.5 次以內,此誤差應是在可容忍的範圍之內。

如將以上之 4 組測試資料其取樣的方法,由以上之按順序取出,改為每次以不同亂數分別取出與以上各組相同數量的資料,分別代入模式運算後,得出其測試之誤差均方根值分別為 0.1191, 0.09973, 0.1193 及 0.1025。而逐年之實際每年侵台颱風次數與本模型計算值,示於圖 6。由此結果可知,測試資料為由亂數取樣取出的,是較使用按順序取樣取出的,其運算誤差結果較佳。

### 4.2. 網路輸入因子與輸出因子之相關性檢定

為檢定各輸入因子與輸出因子(即:每年發生侵台颱風次數)間之相關性,進而比對倒傳遞類神經網路訓練時所產生的敏感度相關性。本研究採用線性相關係數之統計結果來檢定其相關性,其相關係數變化範圍為-1 至+1,當相關係

數趨近+1 時,則兩變數可為完全正相關,當相關係數趨近-1 時,則兩變數可為完全負相關,當其值為 0 時,則兩變數間之關係為無相關,本研究之檢定結果顯示如下:

(1).以 1982 至 2003 年之 22 年資料分析

此區間在本模型之 6 組輸入因子之資料,與輸出因子(即:每年發生侵台颱風次數)資料間之線性相關係數示與於表 1,其與每年發生侵台颱風次數相關性之最大之輸入因子為第 3 組輸入因子(即:前一年之台灣之降雨量),相關係數為負相關,其值為-0.335。

(2).以 1982 至 1992 年之前 11 年資料分析

此區間,其在本模型之 6 組輸入因子資料,與輸出因子(即:每年發生侵台颱風次數)資料間之線性相關係數,示於表 2,其與每年發生侵台颱風次數相關性最大之輸入因子為第 5 組因子(即:今年一月至四月北太平洋東西水域海表面水溫差),其相關係數達-0.353,其次為第 4 組輸入因子(即:前一年北太平洋東西水域海表面水溫差),其相關係數達 -0.348。

(3).以 1993 至 2003 年後 11 年資料分析

此區間,其在本模型之 6 組輸入因子資料,與輸出因子(即:每年發生侵台颱風次數)資料間之線性相關係數示與於表 3,其與每年發生侵台颱風次數相關性最大之輸入因子為第 3 組輸入因子(即:前一年全年全臺灣之雨量),其相關係數達-0.503,其次為第 2 組輸入因子(即:當年 1 月至 4 月之台灣北南雨量差),相關係數達 0.363,及第 1 組輸入因子(即:當年 1 月至 4 月之台灣南北雨量),其相關係數達 0.275。

由此可知,第 1, 2, 3, 4 及第 5 組輸入因子,均與輸出因子,在不同時段,各有其不同程度強烈相關性,亦即本網路之選用輸入因子,有其正確性,然以上結果亦說明:影響輸出因子之主要因子,其前 11 年與後 11 年資料,有相當大的差異,尤其在後 11 年資料中第 3 組輸入資料與輸出資料相關性最大,且為負相關,但在前 11 年資料中它們卻是正相關,這也說明測試資料,如用亂數取樣,其模型運算誤差,是較使用按順序取樣要佳的原因之一。

### 4.3. 2005 年侵臺颱風數的預測

經相關資料輸入本模型後,計算出 2005 年侵臺之颱風數,為 5.7 個。

## 五. 結 論

類神經網路之倒傳遞類神經網路,被運用來計算預測侵台颱風次數之用,因本研究因只有少量資料(共 22 組)可供使用,故交叉驗證法,被用來建模時使用,即將全體資料分批輪流作為測試用的資料,經分析後其可獲的以下主要結論:

1. 本研究建模時,將資料分為 4 組分批輪流作為

測試用的資料，經交叉驗證，其測試之誤差均方根，分別為：0.1234，0.1101，0.0969及0.135，其誤差值在1.5個颱風以內，故本模型為可實際使用的模型。

2. 將6個輸入因子分別與輸出應因子做線性相關性分析，發現各因子在不同時段各有當大的相關性，其相關係數最高可達0.5035之多，也顯示所選用的輸入因子是正確的。
3. 前11年(1983-1993)與後11年(1994-2004)，其影響侵台颱風次數的主要因子是有極大的差異，前11年是與北太平洋東西水域海表面水溫差有極大的相關性，而後11年是與臺灣之雨量有極大的相關性。
4. 經相關資料輸入本模型後，計算出2005年侵臺之颱風數為5.7個。

## 誌謝

本研究承蒙文化大學大氣系余嘉裕主任，張振瑋同學等協助蒐集有關資料，及許志華教授提供本研究有關之建議與指導，在此一併誌謝。

## 參考資料

1. Chu, P. S., and J. D. Clark, (1999) "Decadal Variations of Tropical Cyclone Activity over the Central North Pacific" Bull. Amer. Meteor. Soc., 80:1875-1881。
2. Gray, W. M., (1977) "Tropical Cyclone Genesis in the Western North Pacific", J. of Meteor. Soc, of Japan, 55: 465-481。
3. \_\_\_\_ (1984) "Atlantic Seasonal Hurricane Frequency. Part 1: El Nino and 30mb quasi-biennial Oscillation Influences" Mon. Wea. Rev, 112:1649-1668。
4. Henderson-Sellers, A., H. Zhang, G. Berz, K. Emanuel, w. Gray, C. et al (1998) "Tropical Cyclones and Global Climate Change: A Post-IPPC. Assessment" Bull. Amer. Meteor. Soc., 79(1):19-38。
5. Maloney, E. D. and E. L., Hartmann, (2000) "Modulation of Hurricane Activity in the Gulf of Mexico by the Madden-Julian Oscillation" Sci. 287: 2002-2004。
6. Su, Chao-Ton and Jyh-Hwa Hsu (2005) "An Extended Chi2 Algorithm for Discretization of Real Value Attributes" IEEE Transactions on Knowledge and Data Engineering, 17(3):437-441。
7. Yeh, I. Cheng (2003) "Prediction of Strength of Fly Ash and Slag Concrete by the Use of Artificial Neural Networks" J. of Chinese Institute of Civil and Hydraulic Engineering", 15(4):739-746。
8. 陳婉卿 (2002) "夏季台灣西南部降雨之研究"，中央大學大氣物理研究所碩士論文。
9. 黃威雄 (2000) "類神經網路應用於颱風期間雷達降雨模擬之研究" 台大土木研究所碩士論文。
10. 葉怡成 (2001) 應用類神經網路。儒林書局。台北。
11. 蔡光榮 羅佳明 林成翰 (2004) "類神經網路於921重建區崩塌地自治理工程風險頻評股估模式之建立與應用" 中國土木水利工程學刊。16(2):191-199。
12. 蕭長庚 (1999) "近百年來侵台颱風性質之初步分析報告" 第六屆大氣科學學術研討會論文集:664-667。
13. 蕭裕陸 (2000) "大尺度氣系統變化與西北太平洋颱風的發生頻率與路徑的關係" 台大大氣科學研究所碩士論文。
14. 崔皓盈等 (2003) "聖嬰及反聖嬰年台灣短期氣象要素的變化及所造成的長期影響" 大同高中學生實習氣象站研究報告。

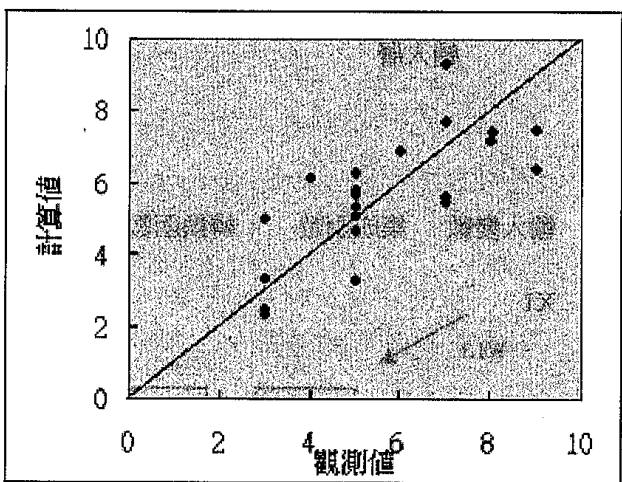


圖 6. 以亂數取樣之實際每年侵台颱風次數與本模型計算值比較圖

表 1. 全體 22 組資料各輸入變數與輸出變數間之相關係數

輸入變數	I	II	III	IV	V	VI
相關係數	0.0865	0.115	-0.335	0.1575	-0.0232	0.0156

表 2. 前 11 年資料各輸入變數與輸出變數間之相關係數

輸入變數	I	II	III	IV	V	VI
相關係數	0.034	0.162	0.184	-0.348	-0.354	0.197

表 3. 後 11 年資料各輸入變數與輸出變數間之相關係數

輸入變數	I	II	III	IV	V	VI
相關係數	0.275	0.363	-0.503	0.2045	-0.223	0.184

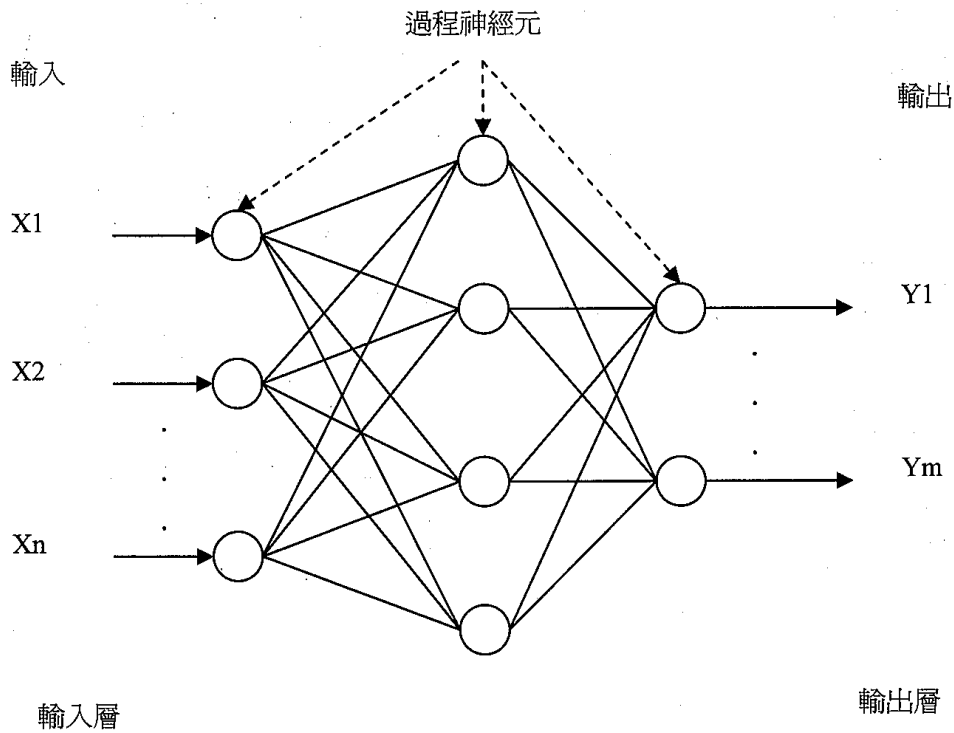


圖 1. 倒傳遞網路架構圖

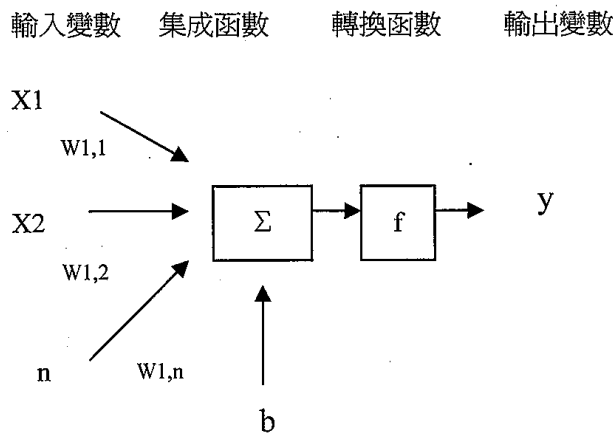


圖 2. 人工神經元之模型

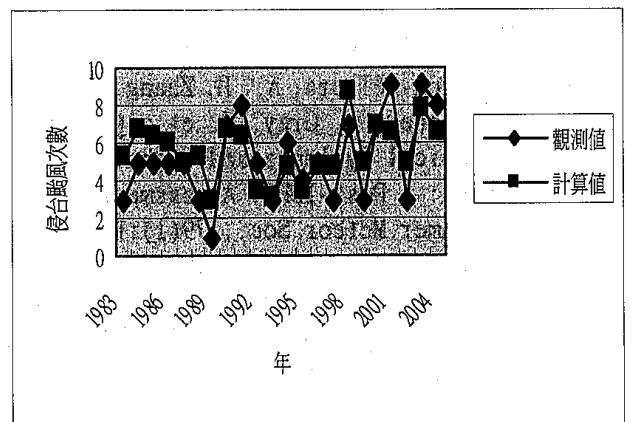


圖 5. 實際每年侵台颱風次數與本模型計算值比較圖

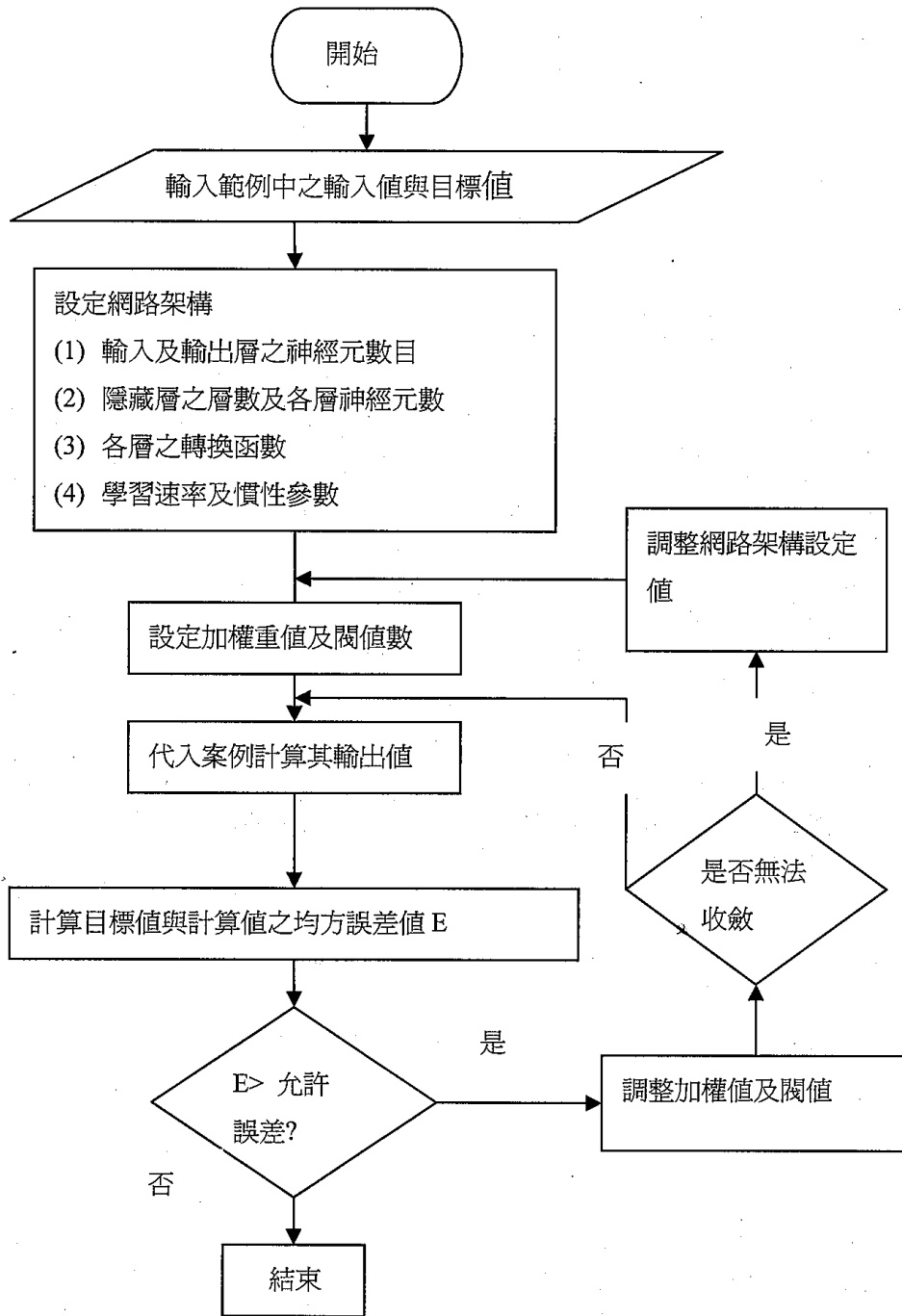


圖 3. 網路學習過程之流程圖

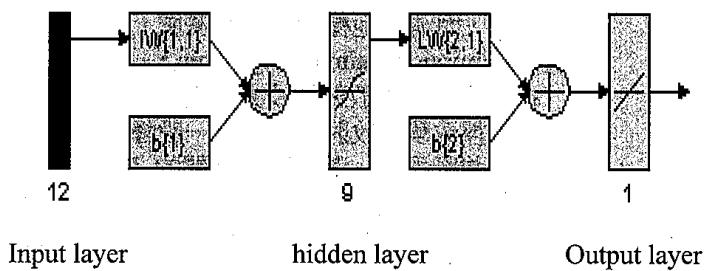


圖 4. 本研究網路架構圖