

國立臺灣大學工學院土木工程學系

碩士論文

Department of Civil Engineering

College of Engineering

National Taiwan University

Master Thesis

應用小波分析辨識地下水水位模擬之類神經網路架構

Structure Identification of Artificial Neural Networks for

Groundwater Simulation Using Wavelet Analysis



Lin, Sheng-Chun

指導教授：徐年盛 教授

Major Professor : Hsu, Nien-Sheng

中華民國 97 年 7 月

Jul. 2008

誌謝

啊，終於畢業了～一不小心就唸了三年= __ =

這篇論文交出去，算是為當初選錯的十多年的升學人生劃下句點，

由衷希望以後不會再和這些有的沒的東西打交道，

可以認真地去做自己真正有興趣的事...：)

最後要謝謝在論文完成過程中給過我幫助的大家，

像我這種腦袋空空的小孩，自己一個人是寫不出啥東西的...（汗）

只不過因為要感謝的人太多，就不一個一個點名感謝了，

以免不小心有人被漏掉了這樣反而不好意思，啊哈哈...|||

總之，不才我畢業了，謝謝大家，大家再見～bye²



摘要

本研究將倒傳遞類神經網路之隱藏神經元個數的決定，視為一般參數檢定的過程，使用 AIC 指標 (Akaike's Information Criterion) 來決定隱藏神經元個數並優選模式。

此外，因為類神經網路的各個權重 w_i 主要是模擬不同生物神經元間的連結強弱，權重為正且越大表示連結越強且增益越大；權重為負且越小表示連結雖強但是為抑制的功用；若權重接近零，表示連結很弱。因此本研究分析在不同隱藏層神經元個數下，各個權重值的大小之變化趨勢，將每一組權重視為一系列的頻譜，找出一適合的門檻值，使用小波分析將之過濾，把過於接近零，也就是連結度越弱的權重予以剔除，進而得出一種更加快速的方式以決定隱藏層神經元個數，以期建立一精確之類神經模式以模擬地下水水位。

結果顯示，以小波分析過濾權重，的確能夠快速地找出合理的隱藏層神經元個數，而且隨著訓練資料筆數的增加，所決定之隱藏神經元個數亦隨之增加，證明了本論文方法論之可行性與可靠度。

關鍵字：類神經網路、小波分析、地下水、AIC 指標

Abstract

Artificial neural network (ANN) is a flexible mathematical structure which is capable of identifying complex nonlinear relationships between input and output data sets. ANN models have been found useful and efficient, particularly in problems for which the characteristics of the processes are difficult to be described using physical equations. Thus, in this paper, the capability of an artificial neural network to provide a data-driven approximation of groundwater flow model is demonstrated.

To identify an ANN model, the number of nodes in every layer must be selected, especially in the hidden layer. In this paper, AIC criterion (Akaike's Information Criterion) is applied to calibrate the ANN model, and the weights of the network are analyzed as a signal by wavelet analysis in order to decide the number of nodes in the hidden layer.

After an mathematical simulation, the result improve that the new approach deduced in this study can quickly offer a useful reference and application for the decision of the number of nodes in the hidden layer.

Keywords: Artificial Neural Networks, Wavelet, Groundwater, AIC criterion

目 錄

誌謝	I
摘要	II
Abstract	III
目錄	IV
圖目錄	VII
表目錄	IX
第一章 緒論	
1.1 研究背景與目的	1
1.2 研究方法概述	2
1.3 類神經網路文獻回顧	3
1.3.1 國外的類神經網路相關論文	3
1.3.2 國內的類神經網路相關論文	4
1.4 小波分析文獻回顧	6
1.4.1 國外的小波分析相關論文	6
1.4.2 國內的小波分析相關論文	6
1.5 論文大綱	7
第二章 小波分析與類神經網路理論介紹	
2.1 前言	8
2.2 類神經網路概述	8
2.2.1 類神經基本介紹	9
2.2.2 類神經網路系統架構	12
2.2.3 類神經網路的學習方式	14
2.3 小波分析概述	15
2.3.1 連續小波轉換	15

2.3.2	離散小波轉換	18
2.3.3	Mallat 運算法	20
2.3.4	小波包	22
第三章 隱藏層神經元個數之決定		
3.1	前言	23
3.2	倒傳遞類神經網路概述	23
3.2.1	BPA 學習演算法	23
3.2.2	BPA 回想演算法	27
3.3	傳統的隱藏層神經元個數決定方式	28
3.4	以 AIC 指標決定隱藏層神經元個數	30
3.4.1	AIC 指標之介紹	30
3.4.2	AIC 指標決定最佳隱藏層神經元個數方法	31
3.5	小波濾除冗餘權重	33
3.5.1	小波濾除冗餘權重理論	33
3.5.2	小波濾除冗餘權重方法	34
第四章 實驗步驟與結果討論		
4.1	前言	36
4.2	實驗流程與詳細步驟	36
4.2.1	均方差與 AIC 之計算	37
4.2.2	小波濾除冗餘權重步驟	47
4.3	AIC 指標方法實驗結果討論	53
4.4	小波濾除冗餘權重結果討論	55
4.5	AIC 指標方法與小波濾除冗餘權重方法之綜合比較	60
第五章 結論與建議		
5.1	研究結論	63

5.2 建議事項	64
參考文獻	65
附錄一 倒傳遞類神經網路公式計算範例	70
附錄二 小波濾除權重範例	75



圖 目 錄

圖 2.1	生物神經元示意圖	9
圖 2.2	類神經網路的基本設計概念	10
圖 2.3	常見的活化函數	12
圖 2.4	前饋式類神經網路	13
圖 2.5	回饋式類神經網路	13
圖 2.6	尺度參數 a 的作用	17
圖 2.7	平移參數 b 的作用	17
圖 2.8	小波的分解與重建示意圖	22
圖 2.9	小波包分解示意圖	22
圖 3.1	倒傳遞類神經網路架構圖	23
圖 3.2	傳統隱藏神經元個數決定範例用圖	29
圖 3.3	以 AIC 指標決定隱藏層神經元個數流程圖	34
圖 3.4	小波濾除冗餘權重流程圖	36
圖 4.1	輸入類神經網路訓練之資料架構	36
圖 4.2	60 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖	43
圖 4.3	60 筆訓練資料下隱藏神經元個數與 AIC 之關係圖	43
圖 4.4	120 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖	43
圖 4.5	120 筆訓練資料下隱藏神經元個數與 AIC 之關係圖	44
圖 4.6	180 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖	44
圖 4.7	180 筆訓練資料下隱藏神經元個數與 AIC 之關係圖	44
圖 4.8	240 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖	45
圖 4.9	240 筆訓練資料下隱藏神經元個數與 AIC 之關係圖	45
圖 4.10	300 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖	45
圖 4.11	300 筆訓練資料下隱藏神經元個數與 AIC 之關係圖	46

圖 4.12	360 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖	46
圖 4.13	360 筆訓練資料下隱藏神經元個數與 AIC 之關係圖	46
圖 4.14	(a)db3 函數圖	48
圖 4.14	(b)墨西哥小帽函數圖	48
圖 4.15	權重排列方式一	48
圖 4.16	權重排列方式二	49
圖 4.17	權重排列方式三	49
圖 4.18	權重排列方式四	49
圖 4.19	權重濾除前示意圖	50
圖 4.20	權重濾除後示意圖	50
圖 4.21	不同資料筆數下最佳隱藏神經元個數示意圖 (小波方法)	52
圖 4.22	180 筆訓練資料下隱藏神經元個數與 AIC 之關係圖 (C=3)	54
圖 4.23	240 筆訓練資料下隱藏神經元個數與 AIC 之關係圖 (C=4)	54
圖 4.24	訓練資料筆數與最佳隱藏神經元數關係圖 (AIC 方法)	54
圖 4.25	訓練資料筆數與最佳隱藏神經元數關係圖 (小波方法, 門檻值=0.6)	58
圖 4.26	AIC 與小波方法結果比較	60
圖 4.27	小波方法不同門檻值對最佳隱藏神經元個數之影響	61

表 目 錄

表 4.1	停止誤差值排序範例	39
表 4.2	60 筆訓練資料下之 AIC 值與計算過程	40
表 4.3	120 筆訓練資料下之 AIC 值與計算過程	40
表 4.4	180 筆訓練資料下之 AIC 值與計算過程	41
表 4.5	240 筆訓練資料下之 AIC 值與計算過程	41
表 4.6	300 筆訓練資料下之 AIC 值與計算過程	42
表 4.7	360 筆訓練資料下之 AIC 值與計算過程	42
表 4.8	小波濾除權重範例	51
表 4.9	不同資料筆數下隱藏神經元個數決定過程	52
表 4.10	訓練資料筆數與最佳隱藏神經元數 (AIC 方法)	54
表 4.11	120 筆訓練資料下權重過濾後神經元減少路徑	55
表 4.12	180 筆訓練資料下權重過濾後神經元減少路徑	56
表 4.13	240 筆訓練資料下權重過濾後神經元減少路徑	56
表 4.14	300 筆訓練資料下權重過濾後神經元減少路徑	57
表 4.15	360 筆訓練資料下權重過濾後神經元減少路徑	57
表 4.16	訓練資料筆數與最佳隱藏神經元數 (小波方法, 門檻值=0.6)	58
表 4.17	不論訓練資料筆數被濾除之隱藏神經元個數 (門檻值=0.6)	57
表 4.18	AIC 與小波方法結果比較	60

第一章 緒論

1.1 研究背景與目的

台灣屬於季風海島型氣候，年平均降雨量約在 2500 公厘左右，比起世界各國來說，水資源算是相當豐沛；但是因為降雨分佈不均，而且山勢坡陡流急，水資源無法有效儲蓄利用，再者近年台灣不論是民生或是工業用水需求量皆大增，便造成了缺水的問題；地上水不足，民眾便轉而抽取地下水，在過量的抽取之後，已經造成海水入侵地下水鹽化、台灣西南沿海地區地層下陷等等嚴重問題；因此如何有效管理並適度使用地下水資源，便成為現今水資源管理的重要課題之一。

欲管理地下水資源，首先必須知道地下水水位的分佈與變化，以便在水位過低的地區禁止抽取並進行補注，在水位尚足的地區適度地抽取利用；傳統欲得地下水水位，通常是透過地下水水流方程式建立地下水水流模擬模式（simulation model）求解，欲求解地下水水流方程式便需要知道方程式裡各個參數的參數值。然而自然界中地下水含水層之變化大而複雜，實測地下水文參數如流通係數、貯蓄係數等，往往因經濟或技術因素之限制，只能侷限於某些特定位置，而這些特定點所測得的地下水參數值未必能代表整個區域的參數型態；要獲得整個分佈式系統之參數實際上是不可可能的，通常的作法是以抽水試驗（pumping test），由洩降曲線反推該抽水井位置附近的流通係數和貯蓄係數值，再利用這些參數將整個區域性地下水系統以參數分區法（zonation method）或內插法（interpolation method）表示成有限維度之方程式，再由參數檢定（parameter identification method）求出參數值，然後將所求之參數應用於地下水水流模擬模式（例如：MODFLOW）中；整個過程繁雜且計算量龐大，進行管理規劃時耗時甚鉅，故本論文之研究目的在於，利用小波分析建構一可靠之倒傳遞類神經網路以供模擬地下水水位，希望能夠解決上述困難。

1.2 研究方法概述

類神經網路 (Artificial Neural Networks, ANNs) 或譯為人工神經網路，其主要的概念是嘗試著模仿人類的神經系統，因為人類的神經系統在語音、聽覺、影像和視覺方面均有很完美的表現，所以期望建立能夠在這些方面也有出色表現的人工模型。過去幾年裡，在不斷進展的研究過程中，已確定類神經網路具有誤差容忍度高 (fault tolerance)、平行運算 (parallel computation) 以及強大的學習能力等特性，可利用於非線性動態系統之鑑別與求解，並且也已經證明人工類神經網路可近似大範圍變化的非線性函數，且能達到所要求的精度並準確到特定條件下；故在不知道現地各種參數的情況下，僅以觀測到的輸入與輸出之成對資料 (input and output data sets) 訓練類神經網路來取代需要大量精準參數與觀測資料方能運作的 MODFLOW 成為新的地下水流模擬模式，理論上是可行的，而且已經有不少前輩做過類似的研究。

小波分析為一門新發展之數學科學，經過十幾年之發展，小波分析在理論及方法上均獲致突破性之進展；小波轉換亦稱為小波分解 (wavelet decomposition)，為一種線性運算，可對訊號進行不同分辨層 (尺度) 之分解，能將各種交織在一起之混合訊號分解成不同分辨層或不同頻率之區塊訊號，可有效地應用於訊號及噪音之分離，提高時頻兩域之分辨能力；它改善了傅立葉轉換 (fourier transformation) 只能觀察到頻率域資料的特性，其二維 (時間和頻率) 分析的功能使得信號在時間和頻率域內的變化能同時被偵測到；故本論文應用小波分析來過濾訓練完成的類神經網路之權重值，觀察其中是否有不必要之噪音存在，以期藉此優選出最佳的隱藏神經元個數，建構一可靠之倒傳遞類神經網路。

1.3 類神經網路文獻回顧

近幾年來，因為理論簡單、結構單純，而且不需使用由複雜的物理或數學理論所建構之物理統計模型，並且容易透過演算法則建立不同條件下所需的特定模式，因此類神經網路大量被應用於各個領域上，在水利相關的研究當然也不少，以下將列舉相關的國內外論文如後。

1.3.1 國外的類神經網路相關論文

French 等 (1992)，將 BPN 運用於空間與時間之降雨強度預測，預測一小時之後的降雨。Daliakopoulos 等 (1994)，使用類神經網路做地下水水位預報，討論不同的類神經網路模式在地下水水位預報上的表現以求得最佳的類神經網路架構所模擬出來之退減曲線。Lorrai 等 (1995)，使用二層隱藏層的 BPN 來架構「降雨—逕流」模式，分析月雨量與月流量間的關係；Hsu 等 (1995)，則使用自己提出的新演算方法 (LLSSIM) 來優化並決定 BPN 架構以探討「降雨—逕流」間的關係，並將結果與兩種傳統的模式 (ARMAX 與 SAC-SMA) 做比較，顯示倒傳遞類神經網路的確能準確模擬「降雨—逕流」間的關係。

Loke 等 (1997)，將類神經網路應用在都市排水的研究上，利用流量歷線或雨量紀錄來預測下一時刻的流量；Asaad 等 (1997)，探討不同的輸入層資料對模擬結果有何影響，其輸入層資料包含雨量、季節性資料、鄰近區域資訊等。

Sajikumar 等 (1997)，將類神經網路應用在月流量的研究上，討論雨量資料缺乏時如何建立模式以推求月流量。Cameron 等 (1997)，探討多種類輸入層資料之應用對月流量之推估有何影響，其輸入層資料包含雨量、溫度及流量。

Zhang 與 Rao (2000)，主要將類神經網路應用於多流域的集流現象，其輸入層資料含雨量及平均溫度；Parios 等 (2000)，以回饋式類神經網路進行多階段的水位預測。Abdalla 與 Grcia (2003)，將類神經網路應用於地下水水流模式之參數檢定，探討地下水水利傳導係數與地下水水頭之間的關係。

1.3.2 國內的類神經網路相關論文

陳昶憲、楊朝仲、王益文（1996），以倒傳遞類神經網路對烏溪流域之洪流預報作研究，以現時及過去 2 小時的資料預測未來 1 小時流量。孫建平（1997），利用 BPN 及 GMDH，取前數小時之流量及雨量資料來預測下一個小時的流量；黃智顯（1997），利用 BPN 研究坡地集水區的特性，取前數個小時的雨量及流量來預測一或多個小時後的流量。

楊朝仲（1997），以類神經網路為模式設計之主體，適時地與線性轉換函數、灰關聯分析及自迴歸移動平均模式作搭配組合使用，設計適用於集水區洪水流量預測之類神經網路，發展出時序類神經洪流預測模式與灰色類神經降雨逕流模式，利用上游的流量資料來預測下游之洪流量。

郭益銘（1998），應用因子分析將雲林沿海地區地下水質資料進行歸納、整理和分類，以評估雲林沿海地區地下水質污染情形，然後應用倒傳遞類神經網路來預測地下水質變化，結果顯示 BP 具有模擬水質複雜變化之能力，並於預測時能獲得良好之可信度。

胡湘帆、黃源義（1998），結合規則庫控制、模糊控制及倒傳遞類神經網路建立一種架構簡單、應用方便，而且具有良好的推估能力及濾波的功能的「模糊類神經網路」，以進行流量推估的工作。

黃群岳（2000），發表之「颱風洪流量之神經網路預測」，以流域內各雨量站不同延時之逐時雨量為輸入值，水庫入流量為輸出值，利用倒傳遞網路建立「降雨—逕流」模式，預測洪峰到達時間及洪峰值。

梁晉銘（2000），彙整並針對多種具代表性之類神經網路模式進行深入的剖析，並逐一歸納其優劣特性，得出若干可行之原則後，結合模糊理論與模糊「最小—最大」團塊分類法對一複合型類神經網路模式（幅狀基底函數類神經網路）進行結構性之改良，建構完成之新模式其結構不但具有建構迅速、組成彈性之優點外，並且能以最簡單的方式快速而有效地完成不同性質水文問題的需求，對於

未來整合河系全渠段不同成因多種水文問題推估系統之建立，提供了進一步實用化之參考。

胡永國 (2002)，針對已於 2001 年開工的高雄捷運主要的紅線路段，利用倒傳遞類神經網路，將現有地層鑽探資料中之土層孔隙比、洩降水位高度以及總體單位重做為輸入參數，分別以三種不同組合模式，來建立各參數與地表壓密沉陷量之相對最佳模式，進而推算出捷運紅線沿線從地表至施工基礎開挖面（地表下 17 至 20 公尺）所可能產生之沉陷量與分佈地區，以作為防治因施工所造成之潛在災害之評估參考。

曾國源 (2001)，以過去幾年曾多次發土石流災害之陳有蘭溪集水區為主要的研究區域，蒐集各類土石流發生資料，配合土石流災害之水文與地文條件進行發生機制辨識，架構一共用臨域類神經網路 (Shared Near Neighbors Network, SNN Network) 作為土石流預警系統之用，並評判其準確度及實用的效果性，藉以對未來可能發生之情況做模擬。

鍾芸菁 (2004)，使用地理資訊系統建置影響因子資料庫，經因素分析選定影響地下水流發生之參數為累積降雨量、降雨強度、主流長度、平均高程等項，隨後採用倒傳遞網路分析影響地下水水位變化因素之相關程度，並決定類神經網路結構之最佳參數，以針對集水區降雨與地下水水位之模擬學習、輸入影響因素個數及加入地文因子學習表現進行探討。

賴建元 (2006)，利用地理統計方法來決定屏東平原地下水位推估之點位數量，再以類神經網路來推估地下水位，發現類神經網路方法於各月份之推估值均能接近實際觀測值，尤以空間推估最為優秀，而且其結果精度比傳統 SURFER 軟體所提供之克利金法佳。

1.4 小波分析文獻回顧

比起類神經網路，目前國內外應用小波分析的水利相關論文其實不多，以下列舉相關的國內外論文如後。

1.4.1 國外的小波分析相關論文

Coulibaly 等 (2000)，應用遞迴神經網路 (recurrent neural networks) 方法及小波分析，探討低頻 (low-frequency) 部分之氣候變化指標，以預測研析區域之年逕流量。Labat 等 (2000)，利用連續小波及離散正交多分辨分析 (discrete orthogonal multiresolution analyses) 於半小時、一小時、日雨量及日流量等不同尺度之記錄資料，探討石灰岩地層 (karstic springs) 之「降雨—逕流」關係。

Masuda 與 Aihara (2001)，應用小波轉換將時間序列表示成不同頻率成分之加總，並應用小波係數進行混沌時間序列 (chaotic time series) 之預測。Bayazit 與 Ahsoy (2001)，將小波視為資料繁衍 (data generation) 之工具，把觀測資料分解成細節，最後藉由隨機加總這些細節以重構並產生新資料來應用於年流量與月流量之預測，以驗證所提方法之可行性。

1.4.2 國內的小波分析相關論文

李宗穆 (1994)，應用小波分析之噪音抽取能力及類神經網路模式之非線性計算能力，建立一種可以同時分析流量資料噪訊及模擬「降雨—逕流」相互關係之時間序列小波串聯模式，用以從事坡地上游集水區「降雨—逕流」歷程之模擬。

王志雄 (1995)，研析感潮河段內「水位—流量」之率定關係，藉著小波理論分解、去噪及重建之功能，串聯合適之水文模式，以建立一種濾除潮汐效應後之修正「水位—流量」率定曲線。

王瀚德(2001)，以水位資料為背景，利用小波函數將水位訊號做多重解析度分析，藉此過濾潮汐中之高頻雜訊部份，並透過分潮延時差的觀念建立一種可以有效預測及補遺水位資料的類神經網路模式。

周建明(2001)，利用 à trous 小波分解方法，將欲研析之水文序列分解成多個分辨層成份之疊加，將水文序列各分辨層輸入離散系統線性黑盒模式與微分擬合之水文灰色模式，以分別模擬各分辨層「降雨—逕流」之歷程。

廖啟佑(2004)，應用類神經網路與小波分析理論，探索地下水位測站之長時程地下水水位之多分辨層結構，將地下水位各種交織在一起之混合訊號分解成不同分辨層或不同頻率區塊訊號，使用小波收縮(wavelet shrinkage)方法選取一合適的臨界值將所得的高頻小波係數做修剪(clipping)收縮處理，並將門檻值以外之各高頻小波係數與地震發生時間做一分析整理，藉此可明顯指出地下水位出現頻率異常之時間點，因而將有助於減災及延長避難反應時間。

由以上的文獻回顧可知，一般來說小波除了很少應用於水利方面之外，應用時亦常與其他科學技術(例如：類神經網路)互相結合；而且通常的作法是先用小波分析序列資料，藉此濾除訊號裡的噪音之後，再將分析後的資料放入類神經網路裡訓練模擬，以提升類神經網路的模擬能力。

1.5 論文大綱

本論文首先以 MODFLOW 產生水位資料，然後訓練類神經網路，最後配合 AIC 指標與小波分析，決定最佳的隱藏神經元個數，各章節安排如下：第一章為緒論，介紹本論文之內容及大綱，與文獻回顧。第二章簡述類神經網路、小波理論。第三章首先簡述本論文的想法與方法論，並且詳細交代本論文中所使用的方法流程。針對本論文所提出的研究方法，以第四章的數值模擬和實作結果來驗證所提出之方法論之有效性。最後第五章述說結論，對本論文作一總結。

第二章 小波分析與類神經網路理論介紹

2.1 前言

本論文以小波分析與類神經網路為主要的研究工具，因此在本章裡將依次介紹類神經網路、小波分析的基本理論。

2.2 類神經網路概述

在1943年，心理學家Warren McCulloch和數學家Walter Pitts共同提出了神經元最早的數學模式（即MP模式），開創了腦神經科學理論研究的時代。西元1949年Hebbian提出了神經細胞的學習規則，其規則描述為「當人腦在學習不同事物時，每個細胞的連結隨時都在改變，如果一個腦細胞受到另一個腦細胞連續的作用時，它們之間的連結力量就會增強」，此學習規則引導了日後類神經網路的發展。在1958年，Frank Rosenblatt首先引用感知器觀念來模擬大腦感知和學習兩大能力；而感知器是由具活化轉移函數的神經元組成的層狀網路，它具有學習功能；但是Marvin Minsky和Seymour Papert合著的「Perceptron」一書指出，由於感知器結構上的限制，它不能產生複雜的邏輯函數，因為Minsky在MIT有崇高學數地位，類神經網路的研究從該書發表後，就一直陷入低潮無所進展。

所幸在1962年Bernard Widrow提出所謂的適應性線性元件（adaline），由於它是一種線性網路，而且具有學習能力，在訊號處理與系統識別等領域受到廣泛的重視和應用。此外，在此挫折時期（1967~1982年），依然有不少學者從事類神經網路方面的研究，為類神經網路以後的發展埋下了堅固的基礎。

到了1980年代初期，著名的物理學家John Hopfield提出了HNN網路，他在此網路中首先引用了能量函數的觀念，為判定網路的穩定性給了依據；Hopfield的研究論文推動了類神經網路的研究，此後，美國國防的DARPA計畫更是大大地推動了類神經網路的研究和進展；當1986年，David Rumelhart和James McClelland共同編著的「Parallel Distributed Processing: Explorations in the Microstructure

of Cognition」一書發表後，更是將類神經網路的研究帶入了新的里程碑，書中所提的倒傳遞網路更是目前最廣為使用的網路；至此，類神經網已進入了蓬勃發展階段。

2.2.1 類神經基本介紹

人類大約有 10^{11} 個神經細胞，每個神經細胞大約有 1000 根連結與其他神經細胞相連結，組成巨大的神經網路系統。生物神經元模型如下圖 2.1。每個生物神經細胞主要由四部分所構成，以下分別說明其功能及其之間的關聯性：

- 1.細胞體 (soma)：是由細胞核、細胞膜、細胞質所組成。
- 2.軸突 (axon)：是一條由細胞體向外延伸的神經纖維 (nerve fiber)，它主要的工作是將由細胞體發射出來的脈波，送到其他的神經元去。
- 3.樹突 (dendrites)：是由細胞體向外延伸的樹狀架構，它主要的功用是被用來接收從其他神經元傳送過來的脈波訊號。
- 4.突觸 (synapse)：神經元之間的訊號是透過軸突和樹突來傳遞，而連接軸突和樹突的界面就是突觸；軸突所傳送出來的脈波訊號會在突觸的地方分泌一種傳導物質，稱為神經傳導物 (neuro transmitter)，這種傳導物是會接收神經元的細胞體，產生電位的變化。

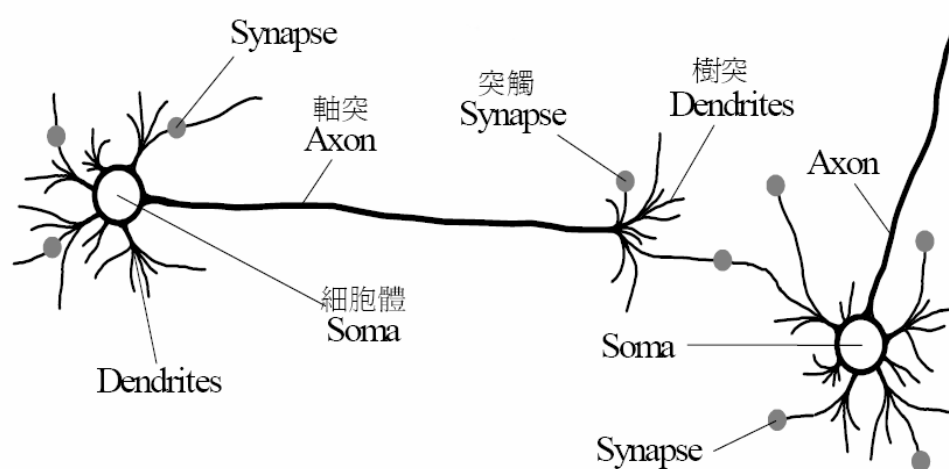


圖 2.1 生物神經元示意圖

人體的神經元是神經系統中最小的訊息及傳遞單元，也是整個神經系統運作的基礎，因此模仿人體神經元的類神經網路，自然係由諸多人工神經元連結而成的網路；人工神經元也可稱為處理單元，圖 2.2 是一個人工神經元的模型，表現類神經網路的基本設計概念，清楚地顯示一個人工神經元的輸入向量 (X)、權重組 (W)、活化函數 (F(·)) 與輸出值 (Y) 的基本關係架構。人工神經元主要分成三個部分：

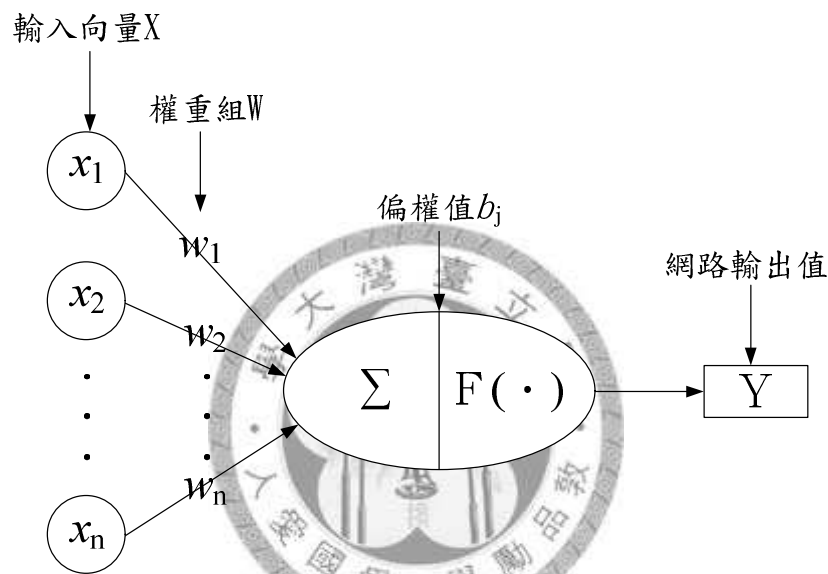


圖 2.2 類神經網路的基本設計概念

1. 一組權重：權重 w_i 主要是模擬不同生物神經元間的連結強弱。權重為正且越大，表示連結越強且增益越大；權重為負且越小，表示連結雖強但是為抑制的功用；若權重接近零，表示連結很弱。
2. 輸入訊號疊加器：主要是模擬生物神經元受多方刺激時膜電位的總變化量；輸入訊號經過不同的權重加權後，在疊加器作線性的疊加。
3. 活化函數：原本是用來模擬生物神經元的門檻值，但目前在運用上已不僅僅用來模擬門檻值，大部分是用來轉化輸入訊息疊加後的輸出值範圍，一般來說經過正規化 (normalized) 的輸出值通常在 $0 \sim 1$ 或是 $-1 \sim 1$ 之間。對一個輸入節點

i 之單一輸入訊號 x_i 而言，可藉由 x_i 與其連結權重 w_{ji} 相乘而將資料傳輸給處理單元 j 。處理單元的工作包含兩部分：第一部分為將其所有傳遞至此神經元的權重信號值做一加總；第二部分的工作則將第一部分之加總值作一非線性的轉換，此部分稱為活化函數。

就數學上而言，一個神經元 j 可以用下列二個方程式來描述：

$$net_j = \sum_{i=1}^m w_{ji}x_i + b_j \quad (2.1)$$

$$y_j = F(net_j) \quad (2.2)$$

式中 y_j 為第 j 個人工神經元輸出訊號，模擬人體細胞的軸突送出訊號； net_j 為輸入值加權值與偏權值 b_j 之總和，模擬人體細胞的膜電位總改變量； $F(net_j)$ 為人工神經元的活化函數，是用來轉換 net_j 的數學函數； w_{ji} 為連結第 i 個輸入值與第 j 個人工神經元之加權值，模擬神經元間的連結強度； x_i 為人工神經元的輸入訊號，模擬樹突傳入訊號； b_j 為偏權值，若為正則對輸入是增益，若為負則抑制輸入值。

常用的活化函數有下列四種：

(1)門檻值函數（圖2.3(a)）

$$F(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases} \quad (2.3)$$

(2)片段線性函數（圖2.3(b)）

$$F(v) = \begin{cases} 1 & , \text{if } v \geq +\frac{1}{2} \\ v + \frac{1}{2} & , \text{if } v > -\frac{1}{2} \\ 0 & , \text{if } v \leq -\frac{1}{2} \end{cases} \quad (2.4)$$

(3)S形函數（圖2.3(c)）

$$F(v) = \frac{1}{1 + \exp(-\alpha v)} \quad (2.5)$$

(4)雙曲線函數 (圖2.3(d))

$$F(v) = \tanh(\alpha v) \quad (2.6)$$

以上四個活化函數的定義範圍都在-1~1之間，可以讓神經元的輸出值維持在合理的範圍內；而活化函數一般為非線性函數，這是因為若採用線性函數，可能會在處理資料時漏失了輸入資料的非線性特性，可能會嚴重影響類神經網路模擬非線性特性的效果，進而對最後的結果造成誤差。

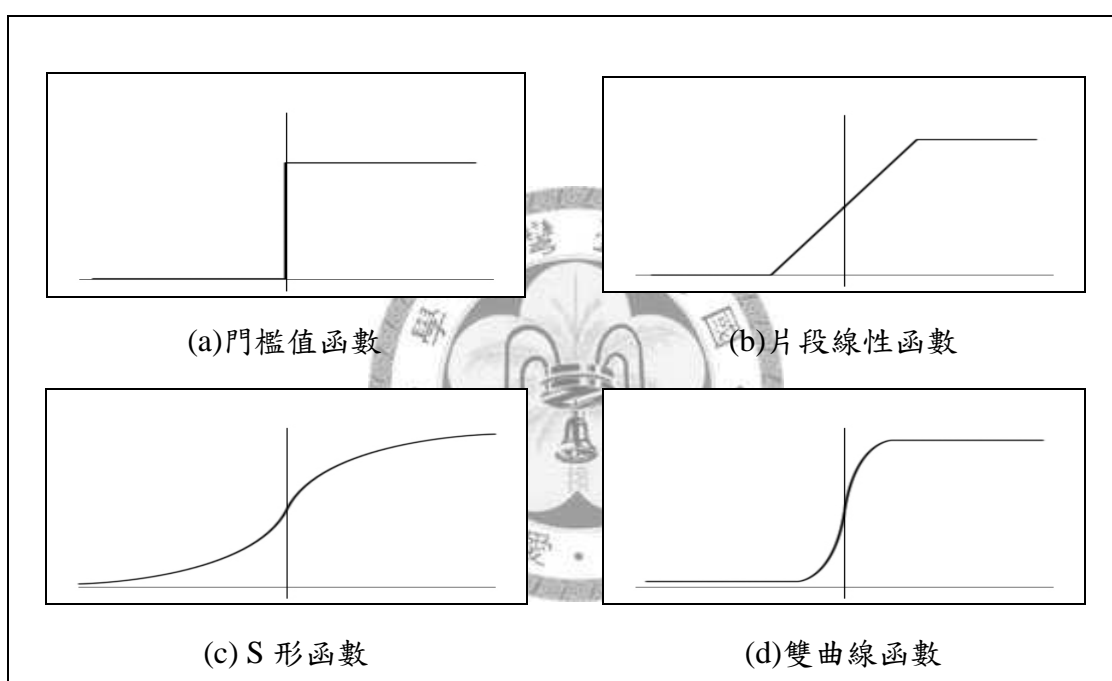


圖 2.3 常見的活化函數

2.2.2 類神經網路系統架構

以連結架構來看，類神經網路可分為前饋式與回饋式兩種：

1. 前饋式類神經網路 (feed-forward)：

前饋式類神經網路之架構，其連結方式為單一方向地向前傳遞連結，且網路的所有神經元皆無後向或是側向的傳遞連結，如圖2.4所示。

2. 回饋式類神經網路 (feed-back Networks)：

回饋式類神經網路與前饋式最大的不同在於，回饋式類神經網路至少會含

有一回饋迴圈，如圖2.5所示，一個回饋式類神經網路可能僅包含一層神經元，而在此層的神經元會各自將其輸出之訊號，回傳給同一層中的其他神經元或前一層中的神經元，以作為輸入資料。

若以網路的層數而言，類神經網路分為單層式 (single-layer) 與多層式 (multi-layer) 兩種；單層式的網路架構只有輸入層 (input layer) 與輸出層 (output layer)，而多層式架構裡除了輸入層與輸出層，至少含有一層隱藏層 (hidden layer)；換言之，單層式與多層式的主要差別就是在於隱藏層之有無，因此若將圖2.4裡的隱藏層拿掉，那它就成為單層的前饋式網路了。

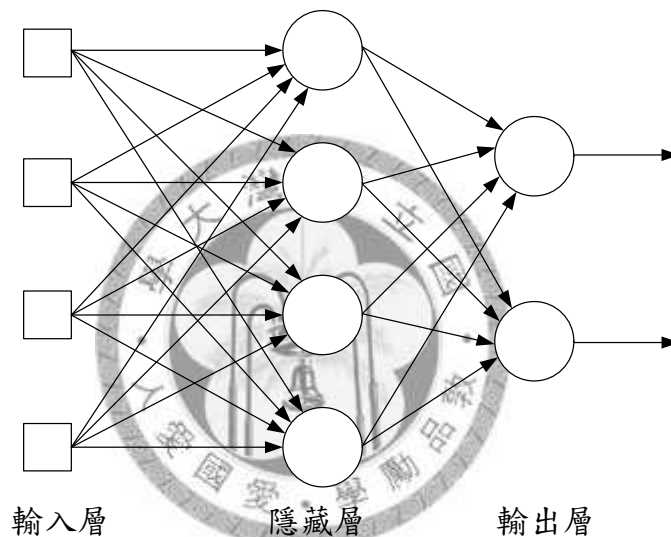


圖2.4 前饋式類神經網路

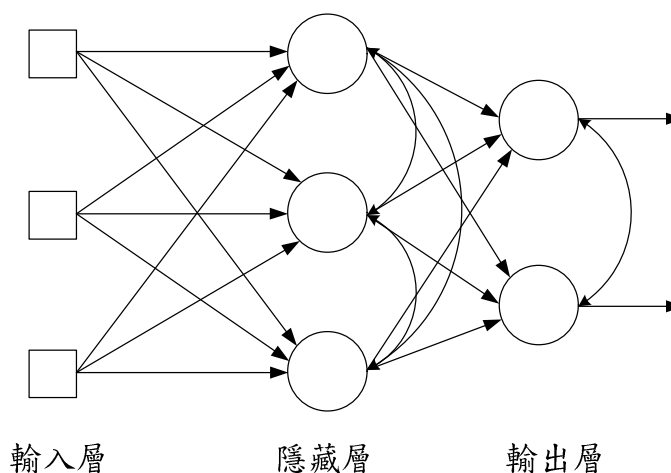


圖2.5 回饋式類神經網路

有一點要在這裡特別提出的是，即便單層式含有輸入與輸出共兩層，但我們仍稱它為「單」層，理由在於輸入層不對訊息作任何處理，只是純粹把資料傳入網路裡，所以不把它視為一層。（當然，也是有其他的學者專家習慣把輸入層視為一層，但本論文不採用這種說法。）

2.2.3 類神經網路的學習方式

類神經網路是以模擬大腦的功能而成，人體大腦的其一特徵即是從經驗中學習，故類神經也有各種不同的學習演算法來模擬這種特性：學習的過程即是調整各人工神經元間的連結權重，而學習的結果（或稱為知識）即為最終所調成出來的權重值大小。此學習的過程大致可分為監督式學習與非監督式學習兩類：

1. 監督式學習（supervised learning）：

監督式學習意即我們監督它學習，以期網路結果可達到我們想要的目標；故在此一學習方式中，我們會給予類神經網路一個訓練範例，範例中同時包含輸入項與監督用的目標輸出值（desired output），然後藉由「學習」不斷修正網路中的連結權重，來降低網路輸出值與目標輸出值間的差距，直到差距小於我們想要的程度才停止。

2. 非監督式學習（unsupervised learning）：

非監督式學習與監督式學習相比較，不同點在於訓練過程中只有提供輸入資料而不提供目標輸出值，讓網路自行依輸入資料的特性自己去學習及調整權重，換言之即是一種聚類（clustering）的過程。

在此學習法裡，類神經網路本身必須從我們輸入的巨量且未經處理的雜亂資料裡，找出任何可能存在的規律性、相關性與個別屬性．．．等等；當類神經網路找出這些特性後，網路本身便會修改其自身的變數以符合資料特性，這一過程稱之為自我組織（self-organization），故非監督式學習適用於真資料之分類特性不確定之時。

2.3 小波分析概述

小波分析的基本理論在於：以數學上函數轉換的方式，將資料分成數個頻率分量；傳統的頻譜分析是以傅利葉轉換 (Fourier transform) 為數學上的理論依據，函數 $f(t)$ 的傅利葉轉換定義為：

$$F(w) = \int_{-\infty}^{\infty} f(t)e^{-iwt} dt \quad (2.7)$$

其中 w 為頻率，所以式(2.7)的意義是將時間函數 $f(t)$ 經過基底函數 e^{-iwt} 投影至各頻率的分量。若以尤拉公式(Euler formula)展開 e^{-iwt} 可得：

$$e^{-iwt} = \cos(wt) - i \sin(wt) \quad (2.8)$$

所以 $F(w)$ 表示的是 $f(t)$ 在整個頻域上的分佈量；而且式(2.7)的時間分佈包括正負無窮大，所以傅利葉轉換缺乏局部時間的解析度。

一般來說，如果訊號的性質並沒有隨時間變化而發生很大變動，則稱這類訊號屬於固定性(stationary)的訊號。但是大多數的情形下，訊號中屬於固定性的分量並非真正希望量測到的，反而是非固定性的分量才是要研究的部分，例如一些不定時產生的突波(spike)或是不規則的漂移。這個時候傅利葉分析在應用上就顯得不夠充分，因為若以傅利葉分析轉換至頻域，將失去時域之相關訊息。

為了修正這個缺點，D. Gabor 提出將訊號切割成許多小段，逐段作傅利葉分析以得到較多時域上的訊息，就是所謂的短時間傅利葉轉換 (Short-Time Fourier Transform, STFT)。但是這個方法也同樣有其限制，就是切段的長度究竟該如何選擇？如果要分析的頻率很高的訊號，當然必須選擇較短的訊號視窗長度 (window length)，如果訊號的頻率很低，則訊號長度必須拉長。但是如果事先無法得知要分析的訊號究竟屬於那個頻帶，則方法也無法使用，小波分析就是為了解決這個問題而推導出來。

2.3.1 連續小波轉換(Continuous Wavelet transform, CWT)

所謂小波，是指一個定義為有限長度且平均值為零的波形，我們可以藉著拉伸 (stretch) 或壓縮 (compress) 小波函數之後，逐段平移 (shifting)，展開被分析的訊號。各式各樣的小波函數，大多延伸自學者 Haar 於西元 1910 年所提出之

Haar 小波函數，所有小波函數皆須滿足 $\int_{-\infty}^{\infty} \psi(t) dt = 0$ 之條件，意指小波函數的平均值必須為零，且定義函數能量有限，故當 x 趨近正負無限大時，小波函數 $\psi(t)$ 將衰減至零。

綜合以上所述，再依據 Rao and Bopardikar (1998) 及 Bayazit and Aksoy (2001) 所定義，小波 $\psi(t)$ 為具有以下二個性質之函數 (Rao and Bopardikar, 1998)：

$$1. \text{函數之積分值為 } 0 : \int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (2.9)$$

$$2. \text{函數之能量有限} : \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \quad (2.10)$$

假設 $f(t)$ 為平方可結合性的函數，則連續小波轉換與小波 (wavelet) 相關之定義如下：

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi \left* \left(\frac{t-b}{a} \right) dt \quad (2.11)$$

其中 CWT 為函數 $f(t)$ 之連續小波轉換值， a 與 b 為實數，分別代表母波 $\psi(t)$ 之尺度參數與平移參數，星號代表共軛複數。 $\psi(t)$ 與 $f(t)$ 是屬於 $L^2(\mathbb{R}^n)$ 空間。在這裡 $L^2(\mathbb{R}^n)$ 是屬於一種信號類別，其包含所有的有限的、定義明確的平方積分函式：

$$f \in L^2 \Rightarrow \int |f(t)|^2 dt = E < \infty \quad (2.12)$$

而這類別簡單代表信號擁有之能量。因此由(2.11)式所知，小波轉換是含有雙變數的函數。小波基底函數為將單一母波函數 $\psi(t)$ 經擴張及轉移後所產生的一組函數，藉由 $\psi_{a,b}(t)$ 可將母波定義為：

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left(\frac{t-b}{a} \right) \quad (2.13)$$

因為 $\psi_{1,0}(t) = \psi(t)$ ，且正規化因子 $\frac{1}{\sqrt{a}}$ 可保證對所有參數 a 與 b 能量維持一定，因此對於任何以給定的值 a ，函數 $\psi_{a,b}(t)$ 只能隨著時間軸將函數 $\psi_{a,0}(t)$ 移動 b 的距離；換言之小波轉換之所以有別於 FFT 及 STFT，即因為它具有此兩個重要的

參數：尺度參數 a (scalar parameter, a) 與平移參數 b (translate parameter, b)；其中尺度參數 a 是將函數圖形作壓縮或伸展，如圖 2.6 所示；而平移參數 b 是將函數圖形在時間軸上作平移，如圖 2.7 所示。

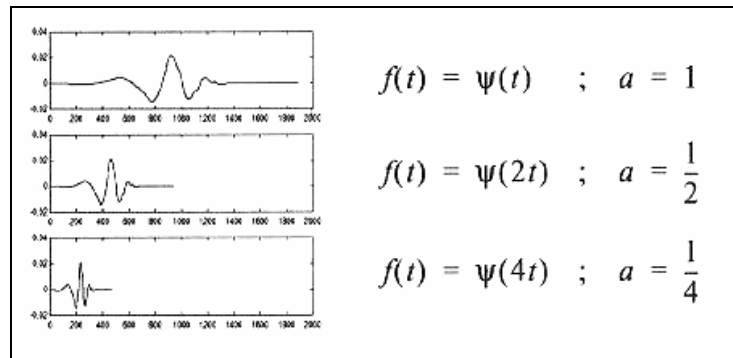


圖 2.6 尺度參數 a 的作用

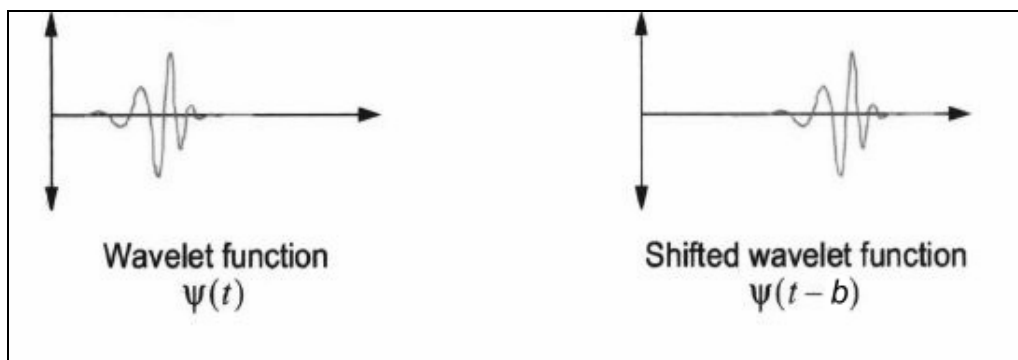


圖 2.7 平移參數 b 的作用

又尺度參數 a 和頻率成倒數，亦即 a 值（尺度）愈小則頻率愈大，圖形被伸展開；而 a 值（尺度）愈大則頻率愈小，圖形被壓縮；尺度參數與頻率間的關係我們可以整理得出如下：

- a 值小(低尺度) \Rightarrow 壓縮小波 (compress wavelet)
 - \Rightarrow 快速變化、細部特徵 (rapidly changing、details features)
 - \Rightarrow 高頻 (high frequency w)
- a 值大(高尺度) \Rightarrow 伸展小波 (stretched wavelet)
 - \Rightarrow 緩慢變化、粗糙特徵 (slowly changing、coarse features)
 - \Rightarrow 低頻 (low frequency w)

這個關係使得小波在頻域中具有縮小 (zoom-in) 及放大 (zoom-out) 的功能。

2.3.2 離散小波轉換(Discrete Wavelet transform, DWT)

由於連續小波轉換受到冗餘性(redundancy)和不可實現性(impracticality)的兩個限制，因此為了兼顧理論的完整性與實用性，可透過(2.5)式對尺度參數 a 和平移參數 b 的取樣而將之離散化(discretization)。對應於上述連續小波轉換函數而言，離散小波轉換可表示如下：

$$DWT(m, n) = \frac{1}{\sqrt{a_0^m}} \sum_k x(k) \psi * \left(\frac{k-n}{a_0^m} \right) \quad (2.14)$$

其中 $DWT(m, n)$ 為信號 $x(t)$ 的離散小波轉換係數值， k 為運算指標，也可以小波級數(wavelet Series)表示之：

$$DWT(m, n) = \frac{1}{\sqrt{2^m}} \sum_k x(k) \psi * \left(\frac{k-n}{2^m} \right) \quad (2.15)$$

比較(2.14)式與(2.15)式可得知，在(2.11)式中的參數 a 與 b 可使用整數參數 m 與 n ，以 $a = a_0^m$ 和 $b = an$ 的方式替換，因此又稱為二的轉換(dyadic transformation)。值得特別注意的是，離散小波轉換仍然是一種連續信號的轉換，其「離散」所指的部分只有在 a 與 b 兩變數而已，時間還是連續的。

Mallat 於 1989 年根據 Laplacian Pyramid Scheme 之架構，推導出多層解析度分析 (multi-resolution analysis) 之基本架構，多層解析度分析主主要是將滿足 $L^2(R^n)$ 空間之任意函數，依連續近似法則作適當分解，分解過程所依據之基底函數 (basis function) 即為小波函數。

依據空間分解 (spatial decomposition) 之觀念，若近似空間 V_j (approximation space) 為依照一定解析度自原始訊號中萃取出來的各種趨勢訊號函數之集合，則可視為低頻訊號之集合，亦即 V_j 之近似訊號 $A_j f$ 表達了訊號函數中最主要的趨勢訊息，當解析度越高 (j 值越大時)，則近似空間 V_j 所能提供的細節訊息就越豐富。於相鄰近似空間 V_{j+1} 與 V_j 對應的兩個趨勢分量 $A_{j+1} f$ 及 $A_j f$ 之差值訊號 $D_j f$ 的集合稱為細節空間 O_j (detail space)。換言之，細節空間 O_j 為近似空間 V_j 在近似空間 V_{j+1} 裡之正交補集，亦即 V_{j+1} 可分解為彼此相互正交的子集合 V_j

與 O_j ，此可以下列數學式表示之：

$$V_j \oplus O_j = V_{j+1} \quad (2.16)$$

$$V_j \cap O_j = \{0\} \quad (2.17)$$

根據(2.8)和(2.9)式，可以得到：

$$L^2(R^n) = O_{-\infty} \oplus \dots \oplus O_{-2} \oplus O_{-1} \oplus O_0 \oplus O_1 \oplus \dots \oplus O_{\infty} \quad (2.18)$$

上式之意義為平方可積之函數空間 $L^2(R^n)$ 是彼此相互正交之差值空間 O_j 的正交直和 (orthogonal direct sum)，這種空間也是一種 Hilbert 空間；根據多重解析度分析，我們可以得到：

$$1. V_j \subset V_{j+1} \quad \forall j \in Z \quad (2.19)$$

亦即在於 $A_{j+1}f$ 具有 A_jf 之所有訊號訊息。

$$2. \lim_{j \rightarrow +\infty} V_j = \bigcup_{-\infty}^{+\infty} V_j = L^2(R^n) \quad (2.20)$$

亦即在於解析度無窮精細時 ($j \rightarrow +\infty$)， A_jf 將趨近於原始訊號 f 。

$$3. \lim_{j \rightarrow -\infty} V_j = \bigcap_{-\infty}^{+\infty} V_j = \{0\} \quad (2.21)$$

亦即在於解析度無窮粗糙時 ($j \rightarrow -\infty$)，信號之訊息將完全遺失。

$$4. g(x) \in V_j, \quad g(x-k) \in V_j, \quad \forall k \in Z \quad (2.22)$$

$$g(x) \in V_j, \quad g(2x) \in V_j, \quad \forall j \in Z$$

$$5. \forall f \in L^2(R^n), \quad \exists \varepsilon > 0 \wedge \exists j \in Z \Rightarrow \|f - A_jf\| < \varepsilon \quad (2.23)$$

$$\text{其中 } \|f - A_jf\| = \sqrt{\int_{-\infty}^{+\infty} (f(x) - A_jf(x))^2 dx}$$

(2.23)式表示對於任一個訊號函數 $f \in L^2(R^n)$ ，吾人均可以找到一個存在於某一近似空間 V_j 內之近似函數 A_jf ，使其近似誤差小於某個給定之有限值。

在多層解析分析裡，吾人定義一個存在於 $L^2(R^n)$ 空間內之標度函數 ϕ ，而且

在每一個近似空間 V_j 定義了正規 (orthonormal) 基底函數 ϕ_{jk} ， $\forall k \in Z$ ；相似地，吾人亦定義了一個存在於每一個細節空間 O_j 中正規化之基底函數 ψ_{jk} ， $\forall k \in Z$ ，稱為小波函數。

多層解析度分析是以存在於近似空間 V_j 內之標度函數 ϕ_{jk} 及細節空間 O_j 內之小波函數 ψ_{jk} 為基礎，藉著存在於彼此具有逐層包含關係的空間 V_j 和 O_j 之正交基底函數，吾人可對 $L^2(R^n)$ 空間內之任意訊號進行反覆多次之分解，將訊號函數在不同頻率層內之分量自原始訊號中萃取出來，達成多層解析度分析；反之，我們亦可以利用 $\phi_{j,k}$ 與 $\psi_{j,k}$ 將 $L^2(R^n)$ 空間內之任意訊號還原，因此 $L^2(R^n)$ 空間內之任意訊號函數 $f(x)$ 亦可利用正交小波函數依(2.15)式表示為：

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} W_{j,k} f_{j,k} \cdot \psi_{j,k}(x) \quad (2.24)$$

上式即稱為小波級數 (wavelet series)，根據上述多層解析度分析的概念，(2.24)式可進一步作如下之分解為：

$$f(x) = A_j f(x) + \sum_{n=j+1}^{\infty} D_n f(x) \quad (2.25)$$

其中 $A_j f(x) = \sum_{n=-\infty}^j \sum_{k=-\infty}^{+\infty} W_{n,k} \psi_{n,k}(x)$ 為訊號函數 $f(x)$ 在近似空間 V_j 內之近似函數；

$D_n f(x) = \sum_{k=-\infty}^{+\infty} W_{n,k} \psi_{n,k}(x)$ 為訊號函數 $f(x)$ 在細節空間 O_n 之分量， $n = j+1 \rightarrow \infty$ ， $n \in Z$ 。

2.3.3 Mallat 運算法

在離散小波分解與重構(wavelet decomposing reconstruction)中，使用Mallat運算法，可將訊號一層層進行分解，每一層分解的結果是將上次分解得到的低頻訊號分解成低頻和高頻兩部分，每一次分解後的數據量減半，因此分析後得到的低頻成分和高頻成分的時域解析度比分解前訊號減低一半。重構算法是分解算法的逆過程，經每一層重構之後，訊號的數據量增加一倍，因此可提高訊號時域解

析度，如圖2.8所示。Mallat分解及重構的過程中，分別需要跳取(downsample or decimate)及零內插(subsample or zero interpolation)，舉例說明如下：

1. 跳取(downsample or decimate)：

假設一訊號序列為[0 1 2 3 4 5 6 7 8]；令跳取因子為2，

則新的序列為[0 2 4 6 8]。

2. 零內插(subsample or zero interpolation)：

經過跳取之後由於序列減半，所以於上述新的序列裡，

依次在每個值之間內插零，即[0 0 2 0 4 0 6 0 8]。

由於小波分解為一個正交分解，因此在訊號分解與重構過程中是不會發生假象現象(aliasing)，當然在訊號取樣的過程必須符合Nyquist取樣定理。

訊號分解是將訊號經由與濾波器摺積(convolution)計算後得到一個新的子空間訊號，濾波器可分為低通濾波器(lowpass filters)及高通濾波器(highpass filters)，低通濾波器由比例函數係數所組合，高通濾波器則是由母小波係數所組合。經由低通濾波器的運算可得到低頻子訊號，經由高通濾波器的運算則可得到高頻子訊號。

訊號重建是將低頻子訊號及高頻子訊號分別經由濾波器的摺積後，將低頻及高頻訊號加起來就可以合成原來的訊號。符號 $\downarrow 2$ 是做次取樣(subsampling)，符號 $\uparrow 2$ 是做升取樣(upsampling)；次取樣是將分解後的訊號用每兩個點取一點的方式重新排序，升取樣是在前面做次取樣後每兩個點之間插入一點零再重新排序。將訊號一層層進行分解，每一層分解的結果為將前次分解得到的低頻訊號分解成低頻及高頻訊號兩部分，且每一次分解後的數據量減半，也因此分析後得到的低頻及高頻成分的時域解析度比分解前之訊號減低一半。

2.3.4 小波包(Wavelet Packet)

小波包是將一組訊號依 Mallat 演算法與小波運算的結果。假設有一組訊號 $S(t)$ ，則可經由 H 濾波器（父小波）與 G 濾波器（母小波）分別分解成低頻的近似訊號（approximation signal） A_1 和高頻的差值訊號（difference signal） D_1 ，然後再一次以 H 濾波器將近似訊號 A_1 分解成下一層的近似訊號 AA_2 與差值訊號 DA_2 ，而 G 濾波器也將差值訊號分解成下一層的近似訊號 AD_2 與差值訊號 DD_2 ，一直重複這個分解過程直到滿足某個條件為止，而這種將訊號分解的過程稱之為小波包分解，它的分解過程如圖 2.9 所示。

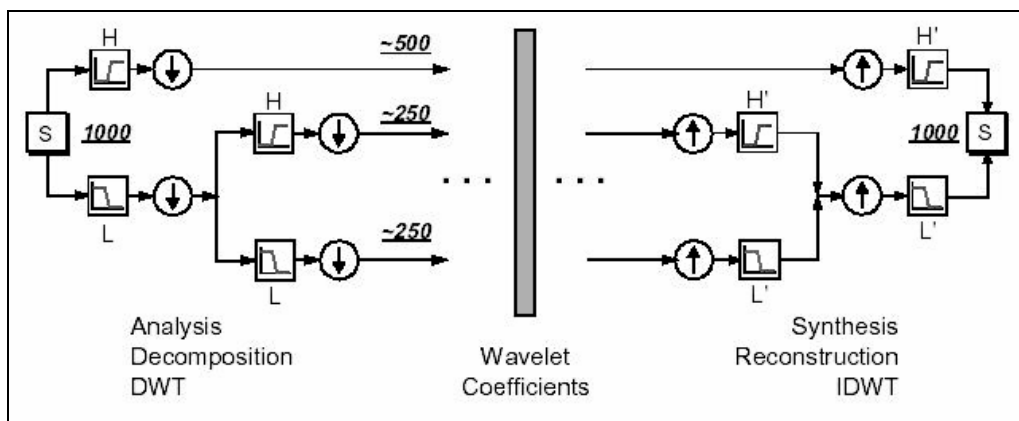


圖 2.8 小波的分解與重建示意圖

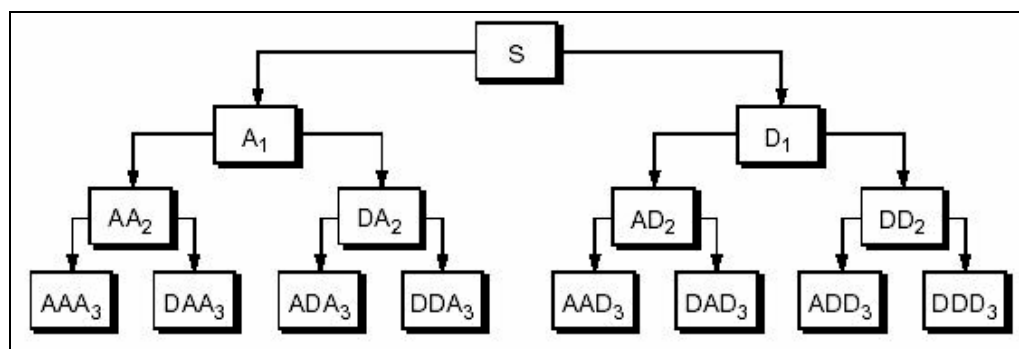


圖 2.9 小波包分解示意圖

第三章 隱藏層神經元個數之決定

3.1 前言

在類神經網路的使用上，除了要選定適當的網路之外，其餘各個參數的決定也是重點，例如隱藏層的層數決定、活化函數的選擇、權重初始化．．．等等，而本研究與本章節則是著重於隱藏層神經元數的決定方式之探討。

3.2 倒傳遞類神經網路概述

倒傳遞類神經網路是最早發明的類神經網路，架構非常簡單，但是精度高而且回想速度快，因此被廣泛地應用，網路架構如圖 3.1 所示；其運作主要分為學習及回想兩部份，學習過程是一種監督式學習，它讀入訓練範例及目標輸出值，將訓練範例輸入至網路中，利用最陡坡降法反覆地調節網路的加權值及偏權值，使得網路輸出值能與目標輸出值更加相近；所謂的回想，即讀入待預測之輸入資料，使用訓練完成的網路模擬然後輸出吾人欲知之值。因為本論文以此種類神經網路為主要分析工具，故分別詳細說明該網路之建構理論與應用步驟於後。

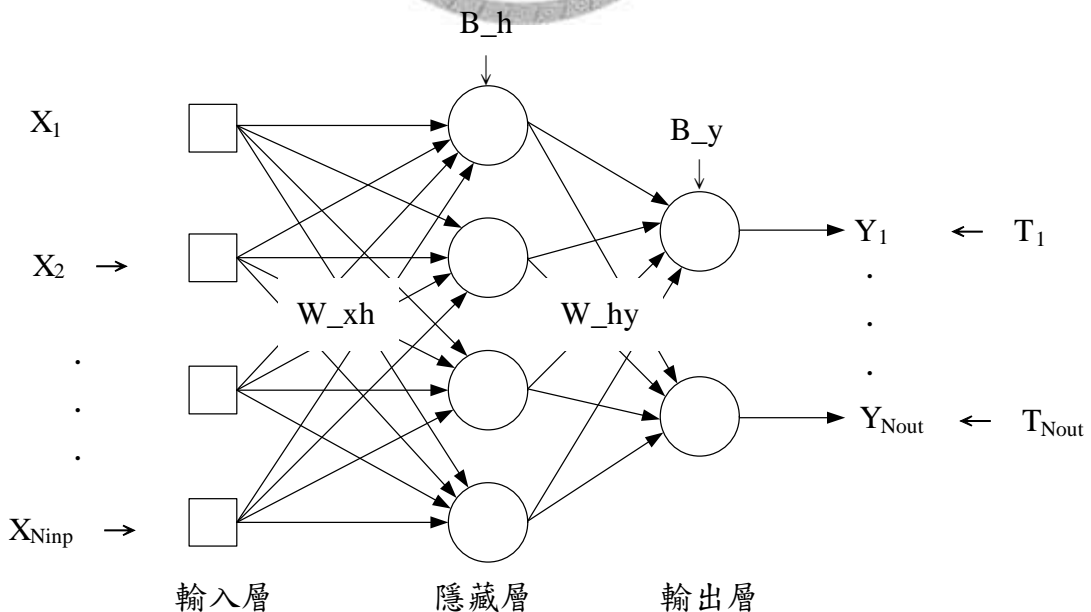


圖 3.1 倒傳遞類神經網路架構圖

3.2.1 BPA 學習演算法

1. 決定網路的層數及各層間神經元數目

一般而言，網路只要具備輸入層，隱藏層一層，和輸出層即可，意即設定網路是三層的架構，令輸入層神經元個數 N_{inp} 個，隱藏層神經元個數 N_{hid} 個，輸出層神經元數目 N_{out} 個；又採用批次學習，令訓練資料 N_{train} 筆。

2. 以均佈隨機亂數設定網路的初始加權值與初始偏權值

不同層的神經元彼此相連，令 $W_{xh}(i,n)$ 為輸入層第 i 個神經元與隱藏層第 n 個神經元間的加權值，由於有 N_{inp} 個輸入神經元與 N_{hid} 個隱藏神經元，故所以可以用一個二層迴圈來設定所有輸入層與隱藏層間的初始加權值如下所示：

```
do n = 1, Nhid, 1
  do i = 1, Ninp, 1
     $W_{xh}(i,n) = \text{random}$ 
```

同理，令 $W_{hy}(n,j)$ 為隱藏層第 n 個神經元與輸出層第 j 個神經元間的加權值，則可以設定所有隱藏層與輸出層間的初始加權值如下：

```
do j = 1, Nout, 1
  do n = 1, Nhid, 1
     $W_{hy}(n,j) = \text{random}$ 
```

再來設定網路的偏權值，只有隱藏層及輸出層才有偏權值，輸入層是沒有的；因為輸入層並沒有运算能力，它只是將一個神經元接收到的訊號平行輸出至隱藏層各個神經元中。令 $B_h(n)$ 為隱藏層第 n 個神經元的偏權值， $B_y(j)$ 為輸出層第 j 個神經元的偏權值，則程式設定如下：

```
do n = 1, Nhid, 1
   $B_h(n) = \text{random}$ 
```

```
do j = 1, Nout, 1
   $B_y(j) = \text{random}$ 
```

3. 輸入訓練樣本 $X[N_{inp}]$ 與目標輸出值 $T[N_{out}]$

4. 計算網路的計算輸出值 $Y[N_{out}]$

首先計算隱藏層的輸出值：

$$\begin{aligned} & \text{do } n = 1, \text{Nhid}, 1 \\ & \text{net_h}(n) = \sum_{i=1}^{\text{Ninp}} W_{_xh}(i, n) \times x(i) + B_{_h}(n) \\ & \text{do } n = 1, \text{Nhid}, 1 \\ & H[n] = \frac{1}{1 + \exp(-\text{net_h}(n))} \end{aligned}$$

其中 $\text{net_h}(n)$ 為隱藏層第 n 個神經元的加權乘積和，而 $H[n]$ 為隱藏層第 n 個神經元的輸出值，它將收集到的加權值乘積和 $\text{net_h}(n)$ 再做一次非線性轉換。

接著，將隱藏層的訊號傳至輸出層：

$$\begin{aligned} & \text{do } j = 1, \text{Nout}, 1 \\ & \text{net_y}(j) = \sum_{n=1}^{\text{Nhid}} W_{_hy}(n, j) \times h(n) + B_{_y}(j) \\ & \text{do } j = 1, \text{Nout}, 1 \\ & Y[j] = \frac{1}{1 + \exp(-\text{net_y}(j))} \end{aligned}$$

其中 $\text{net_y}(j)$ 為輸出層第 j 個神經元的加權乘積和， $Y[j]$ 為網路計算出來的推論輸出值。

5. 計算輸出層、隱藏層與目標結果的差距量

輸出層差距量可如下計算：

$$\text{do } j = 1, \text{Nout}, 1 \\ d_{_y}(j) = Y[j](1 - Y[j])(T[j] - Y[j])$$

其中 $d_{_y}(j)$ 是輸出層第 j 個神經元與目標值的差距量， $(T[j] - Y[j])$ 表示輸出值與目標值間的誤差。

隱藏層差距量的計算如下：

$$\begin{aligned} & \text{do } n = 1, \text{Nhid}, 1 \\ & d_{_h}(n) = H[n](1 - H[n]) \sum_{j=1}^{\text{Nout}} W_{_hy}(n, j) \times d_{_y}(j) \end{aligned}$$

其中 $d_{_h}(n)$ 表示隱藏層第 n 個神經元與目標值的差距量；比較特殊的是上式包

含了一個子項式： $\sum_{j=1}^{\text{Nout}} W_{_hy}(n, j) \times d_{_y}(j)$ ，此式表示輸出層與目標值差距量

的加權乘積和，故 $d_{_h}(n)$ 的計算與輸出層本身的差距量有關，這代表我們將輸

出層的誤差倒傳至隱藏層來計算隱藏層本身的差距量，這也是此種網路「倒傳遞」之稱的由來。

6. 計算各層間的加權值修正量與偏權值修正量

令 $dW_{hy}(n,j)$ 表示隱藏層第 n 個神經元與輸出層第 j 個神經元間的加權值修正量，則計算其間所有加權值修正量的方法如下：

$$\begin{aligned} & \text{do } n = 1, N_{hid,1} \\ & \quad \text{do } j = 1, N_{out,1} \\ & \quad \quad dW_{hy}(n,j) = \eta \times d_y(j) \times h(n) \end{aligned}$$

令 $dB_y(j)$ 為輸出層第 j 個神經元的偏權值修正量，則偏權值修正量的計算方法如下：

$$\begin{aligned} & \text{do } j = 1, N_{out,1} \\ & \quad d_y(j) = -\eta \times d_y(j) \end{aligned}$$

在上面兩個式子中， η 為學習速率，一般取值為 0.1~1.0。有時為了網路的收斂速度，可將上兩式改寫成：

$$\begin{aligned} dW_{hy}(n,j) &= \eta \times d_y(j) \times H(n) + \alpha \times dW_{hy}(n,j) \\ d_y(j) &= -\eta \times d_y(j) + \alpha \times d_y(j) \end{aligned}$$

其中 α 稱為慣性因子，一般取值為 0.0~0.9。

同理，令 $dW_{xh}(i,n)$ 為輸入層第 i 個神經元與隱藏層第 n 個神經元間的加權值修正量，則計算其間所有加權值修正量的方法如下：

$$\begin{aligned} & \text{do } i = 1, N_{inp,1} \\ & \quad \text{do } n = 1, N_{hid,1} \\ & \quad \quad dW_{xh}(i,n) = \eta \times d_h(n) \times x(i) \end{aligned}$$

令 $dB_h(n)$ 為隱藏層第 n 個神經元的偏權值修正量，則偏權值修正量的計算方法如下：

$$\begin{aligned} & \text{do } n = 1, N_{hid,1} \\ & \quad d_h(n) = -\eta \times d_h(n) \end{aligned}$$

7. 更新各層間的加權值及偏權值

$$W_{hy}(n,j)^{t+1} = W_{hy}(n,j)^t + dW_{hy}(n,j)$$

$$W_{xh}(i,n)^{t+1} = W_{xh}(i,n)^t + dW_{xh}(i,n)$$

$$B_h(n)^{t+1} = B_h(n)^t + dB_h(n)$$

$$B_y(j)^{t+1} = B_y(j)^t + dB_y(j)$$

8. 重複步驟 3.~步驟 7.，直至網路收斂

學習過程通常每載入一組範例即更新權重值，但我們亦可採用批次學習，亦即將一批範例載入後，依次求得每一組的權重改變量，將其累加起來再平均以更新權重值（本論文採用批次學習），然後讓網路重複學習數個學習循環，直至網路收斂為止。

為了測試網路是否收斂，定義下列誤差函數來表示網路的學習品質：

$$E = \frac{1}{j} \sum_j (T(j) - Y(j))^2$$

此式表示輸出層各個神經元的平方誤差和；因為在學習過程中，我們希望網路的推論輸出值 $Y(j)$ 與目標輸出值 $T(j)$ 越接近越好，所以亦可在程式中替上式加上一個合理的限制值來做為收斂的條件。

3.2.2 BPA 回想演算法

所謂的回想，即讀入待預測之輸入資料，使用訓練完成的網路模擬然後輸出吾人欲知之值，其過程如下所示：

1. 設定網路的層數及各層間神經元的數目。
2. 讀入已訓練好的網路加權值及偏權值。
3. 讀入已準備好的測試範例 $X(1), X(2), \dots, X(Ninp)$ 。
4. 使用與學習演算法相同的步驟 4.，計算網路輸出值 $Y(1), Y(2), \dots, Y(Ninp)$ ，

隨後便可將結果與真值資料互相比對。

以上即為倒傳遞類神經網路之基礎理論與運作過程，本論文亦以此種類神經網路取代傳統的地下水流模擬模式 (MODFLOW)；而取代過程中必須決定的各種網路參數將逐一詳述如後面章節，唯本研究之重點著重於最佳隱藏神經元個數之決定。

3.3 傳統的隱藏層神經元個數決定方式

倒傳遞類神經網路架構包括輸入層，隱藏層與輸出層，輸入層與輸出層的神經元個數都可以由問題本身決定，例如系統的輸入向量如果有 X 個，那輸入層的神經元個數就是 X 個；如果系統所需的預測變數有 Y 個，那輸出層的神經元個數就是 Y 個。因此當決定隱藏層的層數之後，各個隱藏層神經元的個數多寡對網路將有重大的影響；一旦隱藏層神經元個數過少，類神經網路將沒有足夠的參數來描述 (fit) 訓練範例，反之一旦隱藏層神經元個數過多，將造成參數自由度太高，進而對訓練範例過度地描述 (over-fitting)，甚至將使類神經網路隱含了雜訊的描述而失去歸納推演的能力。是故隱藏層神經元個數之選定，對類神經網路有著極大的影響，以下介紹兩種最常用於決定隱藏層神經元個數的方法：

1. 最傳統的方式之一，乃是使用網路修剪法 (pruning algorithm) 於 1998 年由 Abrahart 等人提出，這個方法是先設定大數量的隱藏層神經元個數開始訓練，再逐一減少神經元個數直到找到適當的隱藏層神經元個數為止，不過這個方法浪費大量的計算時間，比較不經濟。
2. 傳統上的另一方式為使用網路增長法 (constructive algorithm)，是在 1997 年由 Kwok 和 Yeung 提出，此法與網路修剪法剛好相反，是先設定小數目的隱藏層神經元個數開始訓練，再逐一增加神經元個數，直到誤差達容忍範圍為止。下圖 3.2 顯示，使用網路增長法，逐一增加神經元個數，直到均方差 (MSE) 達到容忍範圍或均方差的變化不大時之最少神經元個數，即為最適當的神經元個數，此例中我們可以說最適當的隱藏神經元個數為 5。

由以上討論可知，關於隱藏層神經元個數的決定方式，並沒有快速而且可靠的方法；尤其是傳統的網路增長法與網路修剪法，以逐一增加或是逐一減少的方式觀察權重變化以求得最佳神經元個數，於訓練時間上是相當不經濟的；再者，以圖 3.2 為例，雖說 5 個神經元便以足夠描述訓練範例，但捨棄其他較多的神經元個數又有何根據？這是相當值得探討的問題。

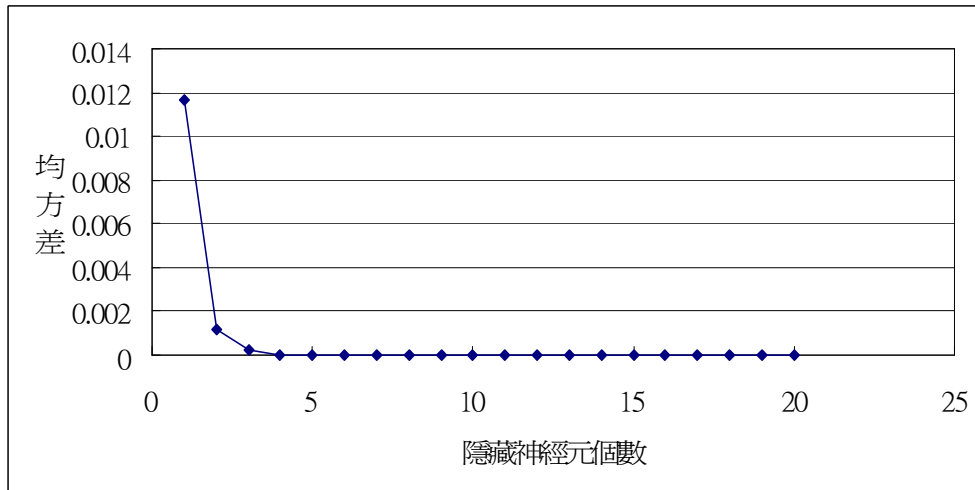


圖 3.2 傳統隱藏神經元個數決定範例用圖

在類神經網路的應用上，會特地談及隱藏層神經元如何決定，或是以新的方法決定的論文頗少見；例如 Abdalla 與 Luis 於 2003 年應用類神經網路於地下水參數檢定時，隱藏神經元 300 個便足以描述整體的模型而且達到誤差收斂標準，然而在此情形下各子分區的精度大小不一，為了將所有的分區精度調到一致，便繼續增加隱藏神經元個數到 500 個為止。又例如 Hsu 等 (1995)，則使用自己提出的新演算方法 (LLSSIM) 來優化並決定 BPN 架構以探討「降雨—逕流」間的關係，並將結果與兩種傳統的模式 (ARMAX 與 SAC-SMA) 做比較，顯示倒傳遞類神經網路的確能準確模擬「降雨—逕流」間的關係。

於是本論文認為，除了依各自實用範例之需要予以變化以決定出最佳的隱藏神經元個數之外，是否亦可將隱藏神經元個數的決定視為一般參數檢定的過程，然後使用其他不錯的參數檢定指標來檢定並且決定隱藏神經元個數？再者如同前面章結對於類神經網路的基本介紹，類神經網路的各個權重 w_i 主要是模擬不同生物神經元間的連結強弱；權重為正且越大，表示連結越強且增益越大；權重為負且越小，表示連結雖強但是為抑制的功用；若權重接近零，表示連結很弱。那麼，我們是否也可以試著分析在不同隱藏層神經元個數下，各個權重值的大小與變化？然後結合參數檢定指標與權重分析兩種方法，進而得出一種更加快速決定隱藏層神經元的方式。

參數檢定指標有很多，本論文決定使用 AIC 指標，採用 AIC 指標的原因在於除了均方差之外，它在模型裡加入考量參數維度與觀測資料數的懲罰項 (penalty term)，以評估在有限觀測下的最佳參數結構，避免傳統僅藉著最大似原原則 (maximum likelihood principle) 來選擇模型，總會選到含有具有最高維度 (參數最多) 的模型。至於權重分析方面，本論文將把每一組權重視為一系列的頻譜，找出一適合的門檻值，使用小波將之過濾分析，把過於接近零，也就是連結度弱的權重予以剔除。

3.4 以 AIC 指標決定隱藏層神經元個數

小波分析的基礎理論已於前一個章節詳細地介紹過，於此不在贅述，這裡將要介紹 AIC (Akaike's Information Criterion) 的發展過程與基礎理論。

3.4.1 AIC 指標之介紹

Akaike Hirotugu (1927-)，名稱漢字寫為「赤池 弘次」，日本的統計學家和數學家，1952 年畢業於東京大學；於西元 1969 年提出著名的 AIC 訊息標準指數 (Akaike's Information Criterion) 用以做為選擇統計模型的準則：

$$AIC = \ln(V) + C \frac{d}{N} \quad (3.1)$$

其中， V 為損失函數 (均方差)；

$d = p + q$ 為自我迴歸移動平均模型之係數的多寡；

N 為適應資料的數量；

C 為介於 2~4 之間的常數值。

又損失函數乃表達資料模型建構的誤差，本研究中採用均方差，而其數值可由下頁 (3.2) 式量測計算得之：

$$V = \frac{1}{N} \sum_{t=1}^N (\varepsilon(x_t, \hat{x}_t) - \hat{\varepsilon})^2 \quad (3.2)$$

其中， N 為離散時間序列之取樣數目；

x_t 為離散時間序列之第 t 個原始取樣資料數值；

\hat{x}_t 為經由自我迴歸移動平均模型產生之第 t 個估計值；

ε 為估計值與原始取樣資料數值之誤差；

$\hat{\varepsilon}$ 為平均誤差值。

3.4.2 AIC 指標決定最佳隱藏神經元個數方法

1. 參考圖 3.3，首先取出現地量測水文資料，包括地下水水位與水文量，並且將之正規化以供類神經網路訓練之用；正規化之時，必須將輸入與輸出向量視為同一向量一起正規化，而非個別正規化，如此方能獲致較佳之訓練結果。
2. 給定一最小之隱藏神經元個數，以 MATLAB 所提供之類神經網路套件模式進行模擬，採用最普通最平常的倒傳遞網路，其架構與設定如下所示：
 - (1) 輸入層一層，隱藏層一層，輸出層一層；
 - (2) 設定網路訓練方法；
 - (3) 設定學習速率與動量係數；
 - (4) 設定訓練次數與訓練停止條件（MSE 達要求精度）；
 - (5) 設定初始權重與初始偏權值。
3. 取出事先準備好的資料使用上述的類神經網路開始訓練，記錄訓練停止後的誤差值，計算對應之 AIC 值。
4. 觀察此 AIC 值是否最小，若是，則此時之隱藏神經元個數即為最佳；若否，則增加神經元個數，重覆步驟 3. 直到 AIC 值最小為止。
5. 改變訓練資料筆數，重覆步驟 3. 與步驟 4.，以觀察不同訓練資料筆數與最佳隱藏神經元個數之關係

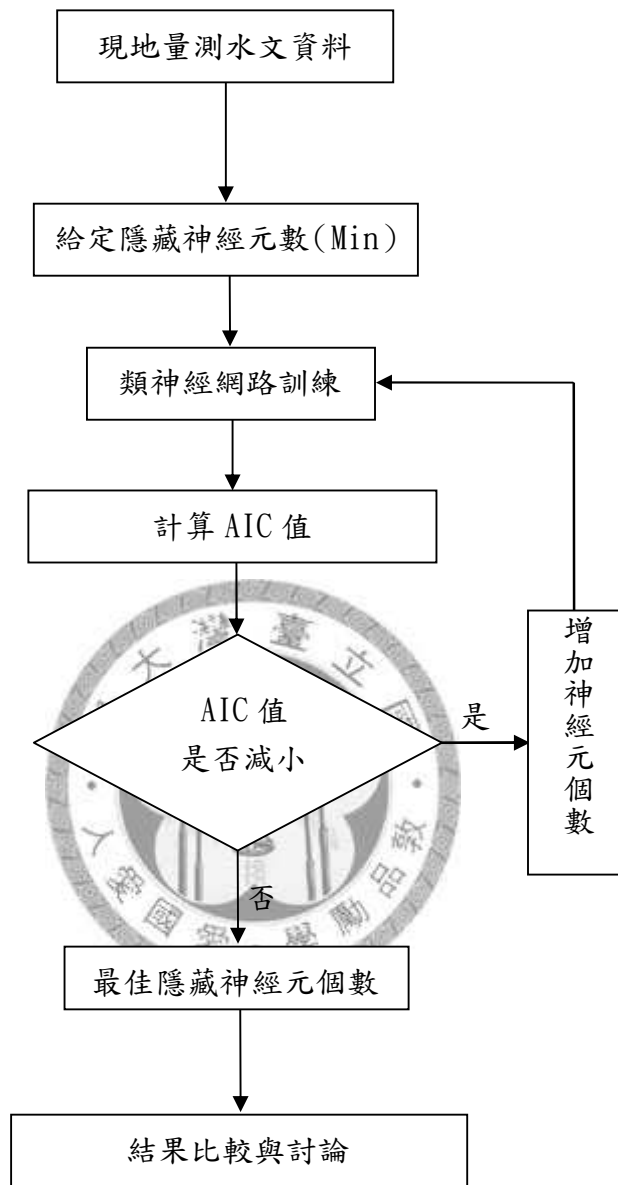


圖 3.3 以 AIC 指標決定隱藏層神經元個數流程圖

3.5 小波濾除冗餘權重

前一章節為 AIC 指標決定最佳隱藏神經元個數方法之介紹，於本章節將提出本論文的重點，以小波分析濾除冗餘權重之想法與實際之作法；關於小波分析之基礎理論，於第二章已詳細介紹，於此不再贅述。

3.5.1 小波濾除冗餘權重理論

本研究擬利用小波分析類神經之神經元連結權重，將其視為一個訊號序列，利用小波轉換將原始訊號分解為平均值 (A1) 與擾動項 (D1)，平均值代表訊號的平均大小，擾動項則表示局部訊號偏離整體趨勢的程度。如果擾動項的值大於所給定之門檻值 (S1)，表示此一訊號偏離平均值甚遠，所以相對而言這個局部訊號相較於鄰近的訊號是個較強的訊號，即較重要的神經元連結，所以將此一連結予以保留，其餘擾動項值較小的連結權重為相對不重要的連結，可將其捨棄，門檻值運作之標準如下式所示：

$$VIP(i) = \begin{cases} 0 & , |D1(i)| < S1 \\ xx(i) & , |D1(i)| \geq S1 \ \& \ |D1(i)/xx(i)| \geq 0.1 \end{cases} \quad (3.3)$$

其中 $VIP(i)$ 代表被保留之原始訊號；

$xx(i)$ 代表第 i 個原始訊號；

$D1(i)$ 代表第 i 個擾動項；

$S1$ 即為本論文所謂之門檻值。

如此一來，透過小波可以濾除不重要的神經元，保留真正重要的神經元；而且透過小波分析不必再以窮舉法慢慢減少神經元數，而是直接可以大幅刪減冗餘的神經元，以快速得到最佳的神經元設置。

3.5.2 小波濾除冗餘權重方法

1. 參考圖 3.4，首先取出現地量測水文資料，包括地下水水位與水文學量，並且將之正規化以供類神經網路訓練之用；正規化之時，必須將輸入與輸出向量視為同一向量一起正規化，而非個別正規化，如此方能獲致較佳之訓練結果。
2. 給定一最大之隱藏神經元個數，以 MATLAB 所提供之類神經網路套件模式進行模擬，採用最普通最平常的倒傳遞網路，其架構與設定如下所示：
 - (1) 輸入層一層，隱藏層一層，輸出層一層；
 - (2) 設定網路訓練方法；
 - (3) 設定學習速率與動量係數；
 - (4) 設定訓練次數與訓練停止條件（MSE 達要求精度）；
 - (5) 設定初始權重與初始偏權值。
3. 取出事先準備好的資料使用上述的類神經網路開始訓練，紀錄訓練停止後的誤差值，還有輸入層與輸出層的權重值。
4. 將權重值放入小波模式裡過濾，觀察是否有隱藏神經元被濾除；若有，則減少神經元個數，回到步驟 3. 重新訓練模式；若無，則以此神經元個數為最佳之隱藏神經元個數。
5. 改變訓練資料筆數，重覆步驟 3. 與步驟 4.，以觀察不同訓練資料筆數與最佳隱藏神經元個數之關係。

以上即為本論文方法論之概述與流程；在下一個章節裡，除了會更加詳細地解說實驗流程之外，亦會交代模式中各個參數的設定值與為何選擇這個參數值，同時也會解說何謂「以小波濾除冗餘權重」之作法；最後再針對本論的實驗結果做一綜合性的比較與探討。

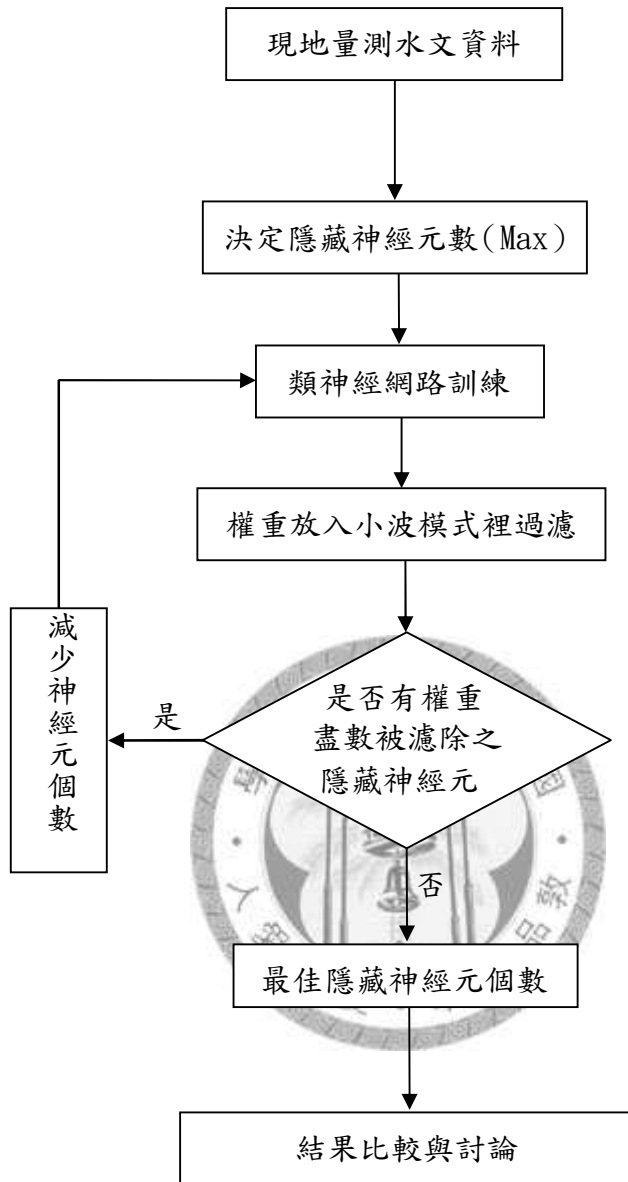


圖 3.4 小波濾除冗餘權重流程圖

第四章 實驗與結果討論

4.1 前言

在前一章節之中已經詳細交代本論文之方法論，本章將更進一步詳述實驗流程與步驟，以及各參數的詳細設定，最後將實驗所得的結果進行討論與比較。

4.2 實驗流程與詳細步驟

本實驗主要流程在於先讀入現地量測之水文資料以供類神經網路訓練，接著分別使用 AIC 指標與小波分析，決定最佳的隱藏神經元個數，最後針對兩種方法之結果予以比較與討論；然現地資料量測不易且有所缺漏，再者本研究之方法論尚在雛形階段，故水文資料暫以 MODFLOW 產生者代替，共產生 360 筆資料，每筆資料包含 8 個輸入向量與 4 個輸出輸出向量，如下圖 4.1 所示：

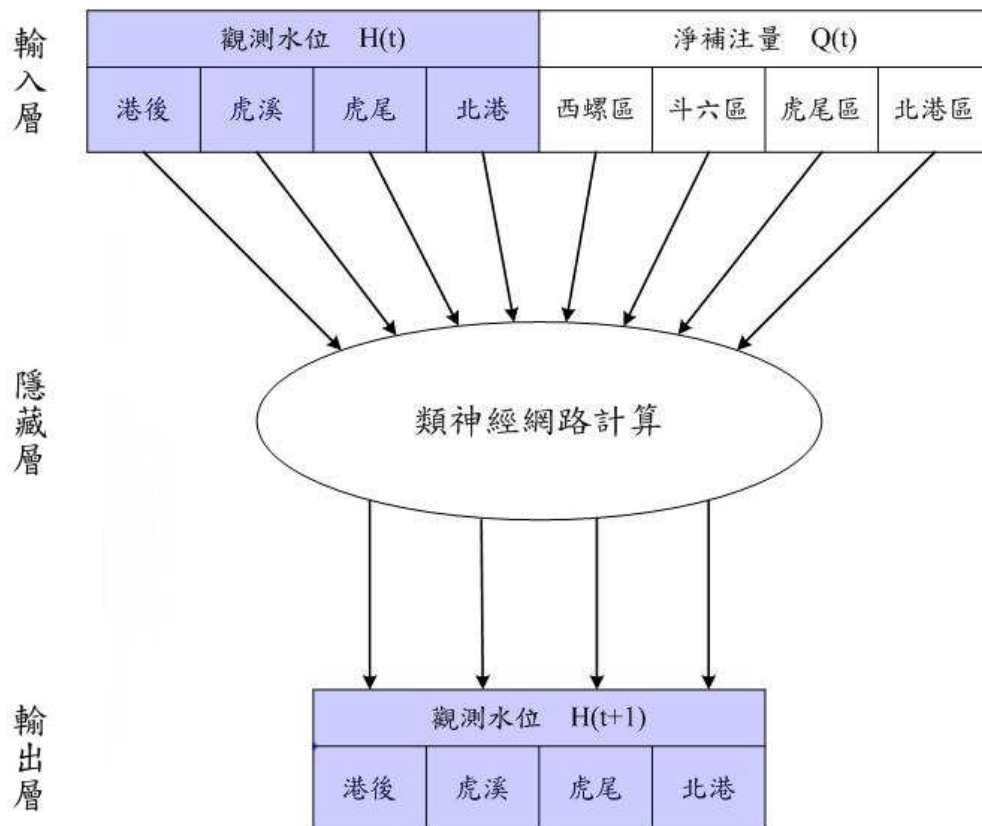


圖 4.1 輸入類神經網路訓練之資料架構

唯因為當初論文之重點著重於訓練類神經網路產生權重以供分析，故輸入 MODFLOW 裡的詳細參數值一時疏忽未加以保留，皆已佚失；目前唯一可知者，觀測水位資料單位為「公尺/旬」，淨補注量單位為「千噸/旬」；因此本研究以目前所做之數值實驗資料也無法再深入探討不同 Time Step 對本論文方法之影響，此部份或許可留待後人研究，本論文將不再補上。

4.2.1 均方差與 AIC 計算步驟

1. 首先以 MODFLOW 產生 360 筆資料，並且將之正規化，以供類神經網路訓練之用，每筆資料包含 8 個輸入向量與 4 個輸出輸出向量。

2. 以 MATLAB 程式設計類神經網路，使用最普通最平常的倒傳遞網路，各參數的設定如下所示：

- (1) 隱藏層一層，活化函數使用 tansig，為雙彎曲正切函數；
- (2) 輸出層活化函數使用 purelin，為線性函數；
- (3) 訓練法使用 trainrp，稱為 Resilient Backpropagation 學習法；
- (4) 學習速率設為 0.001，動量係數為 0.9；
- (5) 訓練次數為 50000 次，訓練目標精度 $MSE=0.00001$ (如果 MSE 未達 0.00001，而訓練次數先滿 50000，訓練仍然會停止)
- (6) 初始權重與初始偏權值皆為 random。

3. 取出事先準備好的資料的前 60 筆，使用上述的類神經網路開始訓練，訓練過程由隱藏神經元個數為 1 個（極少）開始訓練，相關過程描述如下：

- (1) 因為倒傳遞類神經網路之結果為多值解，隨機的初始權重亦可能影響最後停止誤差之大小，為了能夠得到各個不同隱藏層神經元個數下的最佳停止誤差值與最佳權重值，所以每個神經元個數之訓練皆循環 5 次，取 MSE 最小的該次進行分析。
- (2) 參考表 4.1，此為訓練資料筆數 60 筆的情形下，隱藏神經元個數 1~7 的五次的停止誤差值；表左方為未排序，表右方為排序後的結果，由表可知在

1 個隱藏神經元的情形下，停止誤差值取第五次，2 個隱藏神經元的情形下，停止誤差值取第一次，餘類推。

4.逐一增加隱藏神經元個數，依前節所述之 AIC 公式，設定 C 值為 2，計算出對應之 AIC 值，其過程與結果如表 4.2 所示；接著以停止誤差值（MSE）為 Y 軸，隱藏神經元個數為 X 軸作圖，得到 60 筆訓練資料下的隱藏神經元個數與停止誤差值之關係圖，如圖 4.2 所示；再以 AIC 為 Y 軸，隱藏神經元個數為 X 軸作另張圖，得到 60 筆訓練資料下的隱藏神經元個數與 AIC 之關係圖，如圖 4.3 所示。相關細節說明如下：

(1)在本研究之範例中，C 值大小對於 AIC 之影響只在整體 AIC 值之大小或曲線「翹起」之幅度，對於最重要的最小值之落點其實並無實際影響，針對此點於 4.2 節會有所討論。

(2)在本研究之範例中， $d+q$ （請參考前章之(3.1)式）之值僅取類神經網路之連結權重個數加總，並未將偏權值之個數併入計算；原因在於本論文分析之重點為各個連結權重，而且比起網路的連結權重個數，本研究之範例中的偏權值個數相對較少，經過實際計算證明並不會對 AIC 指標之應用結果產生影響。

5.重覆步驟 3.與步驟 4.，逐一增加隱藏神經元個數，一般來說直到 AIC 值不再減小反而增加時即可停止，但因為本研究範例中類神經最後之停止誤差值在隱藏神經元到達 4 個之後，雖然繼續增加隱藏神經元個數，均方差之精度皆收斂在相同精度之下不再有大變動，唯受到初始權重隨機之影響，訓練停止後之均方差值有略為大小起伏之現象，故本研究最後決定將隱藏神經元個數一直增加到 20 個，然後取 AIC 最小值對應之神經元個數為最佳隱藏神經元個數。

6.依次增加訓練資料筆數為 120、180、240、300、360 筆，重覆步驟 3.~5.，便可得到在各個訓練資料筆數下的隱藏神經元個數與 AIC 值列表（如表 4.3~表 4.7 所示），隱藏神經元個數與停止誤差值之關係圖，以及隱藏神經元個數與 AIC 之關係圖，分別如圖 4.4~圖 4.13 所示。

表 4.1 停止誤差值排序範例

神經元數	次數	TRV_60		神經元數	次數	TRV_60
1	1	0.011628114		1	5	0.011627805
	2	0.011627895			2	0.011627895
	3	0.011628354			1	0.011628114
	4	0.011628376			3	0.011628354
	5	0.011627805			4	0.011628376
2	1	0.001195068		2	1	0.001195068
	2	0.001199638			4	0.001197881
	3	0.001199578			3	0.001199578
	4	0.001197881			2	0.001199638
	5	0.001200234			5	0.001200234
3	1	0.000252425		3	1	0.000252425
	2	0.000257273			4	0.000253349
	3	0.000258979			5	0.000254288
	4	0.000253349			2	0.000257273
	5	0.000254288			3	0.000258979
4	1	4.36859E-06		4	1	4.36859E-06
	2	5.34071E-06			2	5.34071E-06
	3	1.01645E-05			5	6.73183E-06
	4	1.96415E-05			3	1.01645E-05
	5	6.73183E-06			4	1.96415E-05
5	1	8.7342E-06		5	4	3.46678E-06
	2	9.6863E-06			1	8.7342E-06
	3	1.12897E-05			2	9.6863E-06
	4	3.46678E-06			3	1.12897E-05
	5	1.89002E-05			5	1.89002E-05
6	1	1.06276E-05		6	4	6.09223E-06
	2	7.19445E-06			2	7.19445E-06
	3	7.47181E-06			5	7.30268E-06
	4	6.09223E-06			3	7.47181E-06
	5	7.30268E-06			1	1.06276E-05
7	1	6.34611E-06		7	4	4.1197E-06
	2	5.98317E-06			2	5.98317E-06
	3	6.90095E-06			1	6.34611E-06
	4	4.1197E-06			3	6.90095E-06
	5	9.51192E-06			5	9.51192E-06

表 4.2 60 筆訓練資料下之 AIC 值與計算過程

TRV_60	V	lnV	d	C*d/N	AIC
1	0.011627805	-4.45435604	12	0.4	-4.05435604
2	0.001195068	-6.729552141	24	0.8	-5.929552141
3	0.000252425	-8.284395512	36	1.2	-7.084395512
4	4.36859E-06	-12.34107026	48	1.6	-10.74107026
5	3.46678E-06	-12.57228435	60	2	-10.57228435
6	6.09223E-06	-12.00849637	72	2.4	-9.608496369
7	4.1197E-06	-12.39973021	84	2.8	-9.599730213
8	2.2271E-06	-13.01481027	96	3.2	-9.814810267
9	2.36683E-06	-12.95395905	108	3.6	-9.353959051
10	2.08907E-06	-13.07879157	120	4	-9.078791567
11	1.56121E-06	-13.3700494	132	4.4	-8.970049396
12	2.96589E-06	-12.7283334	144	4.8	-7.928333402
13	2.19161E-06	-13.03087412	156	5.2	-7.830874124
14	2.27997E-06	-12.99134827	168	5.6	-7.391348273
15	2.02324E-06	-13.11081037	180	6	-7.110810371
16	2.29001E-06	-12.98695437	192	6.4	-6.586954374
17	2.6806E-06	-12.82946991	204	6.8	-6.029469908
18	2.4191E-06	-12.93211499	216	7.2	-5.732114988
19	2.55458E-06	-12.87762273	228	7.6	-5.277622731
20	2.83614E-06	-12.77306659	240	8	-4.773066585

表 4.3 120 筆訓練資料下之 AIC 值與計算過程

TRV_120	V	lnV	d	C*d/N	AIC
1	0.100242508	-2.300162953	12	0.2	-2.100162953
2	0.006798928	-4.990990319	24	0.4	-4.590990319
3	0.001021974	-6.886019404	36	0.6	-6.286019404
4	5.66998E-06	-12.08032497	48	0.8	-11.28032497
5	6.46369E-06	-11.9493102	60	1	-10.9493102
6	7.86531E-06	-11.75304861	72	1.2	-10.55304861
7	5.07347E-06	-12.19148556	84	1.4	-10.79148556
8	7.09906E-06	-11.85554818	96	1.6	-10.25554818
9	7.03298E-06	-11.86490004	108	1.8	-10.06490004
10	6.42782E-06	-11.95487511	120	2	-9.954875113
11	4.60423E-06	-12.28853511	132	2.2	-10.08853511
12	6.17586E-06	-11.99486241	144	2.4	-9.594862414
13	0.000005725	-12.07066801	156	2.6	-9.470668009
14	3.90131E-06	-12.45419816	168	2.8	-9.654198164
15	3.95314E-06	-12.44100036	180	3	-9.441000358
16	3.59152E-06	-12.53693505	192	3.2	-9.336935047
17	3.96462E-06	-12.43810055	204	3.4	-9.038100546
18	4.06404E-06	-12.41333301	216	3.6	-8.813333005
19	4.02737E-06	-12.422397	228	3.8	-8.622397001
20	4.20478E-06	-12.37928858	240	4	-8.379288585

表 4.4 180 筆訓練資料下之 AIC 值與計算過程

TRV_180	V	lnV	d	C*d/N	AIC
1	0.092338206	-2.382297291	12	0.133333	-2.248963958
2	0.04879884	-3.020048743	24	0.266667	-2.753382076
3	0.000455638	-7.693812582	36	0.4	-7.293812582
4	1.69465E-05	-10.98544865	48	0.533333	-10.45211531
5	1.07396E-05	-11.44157085	60	0.666667	-10.77490418
6	9.31557E-06	-11.58382336	72	0.8	-10.78382336
7	6.0255E-06	-12.01951009	84	0.933333	-11.08617676
8	6.4458E-06	-11.9520818	96	1.066667	-10.88541514
9	8.31589E-06	-11.69734242	108	1.2	-10.49734242
10	4.71892E-06	-12.2639306	120	1.333333	-10.93059726
11	0.000005154	-12.17573745	132	1.466667	-10.70907078
12	7.11494E-06	-11.85331376	144	1.6	-10.25331376
13	6.66548E-06	-11.91856859	156	1.733333	-10.18523526
14	4.60375E-06	-12.28863937	168	1.866667	-10.4219727
15	5.29486E-06	-12.14877402	180	2	-10.14877402
16	7.05672E-06	-11.8615302	192	2.133333	-9.72819687
17	5.37798E-06	-12.13319772	204	2.266667	-9.866531052
18	5.73286E-06	-12.06929602	216	2.4	-9.669296024
19	4.49037E-06	-12.31357545	228	2.533333	-9.780242121
20	5.68785E-06	-12.07717824	240	2.666667	-9.41051157

表 4.5 240 筆訓練資料下之 AIC 值與計算過程

TRV_240	V	lnV	d	C*d/N	AIC
1	0.099851398	-2.304072219	12	0.1	-2.204072219
2	0.066715032	-2.707324987	24	0.2	-2.507324987
3	0.000393564	-7.840267801	36	0.3	-7.540267801
4	1.93873E-05	-10.85089441	48	0.4	-10.45089441
5	1.03042E-05	-11.48295898	60	0.5	-10.98295898
6	9.95213E-06	-11.51772396	72	0.6	-10.91772396
7	8.50624E-06	-11.67471055	84	0.7	-10.97471055
8	6.47332E-06	-11.94782144	96	0.8	-11.14782144
9	6.28535E-06	-11.97728903	108	0.9	-11.07728903
10	8.52652E-06	-11.67232925	120	1	-10.67232925
11	5.13061E-06	-12.180286	132	1.1	-11.080286
12	7.05747E-06	-11.86142393	144	1.2	-10.66142393
13	7.21847E-06	-11.83886754	156	1.3	-10.53886754
14	4.92604E-06	-12.22097514	168	1.4	-10.82097514
15	6.01028E-06	-12.02203922	180	1.5	-10.52203922
16	4.89072E-06	-12.22817103	192	1.6	-10.62817103
17	5.52381E-06	-12.10644272	204	1.7	-10.40644272
18	3.85318E-06	-12.46661178	216	1.8	-10.66661178
19	4.53606E-06	-12.30345176	228	1.9	-10.40345176
20	4.30051E-06	-12.35677694	240	2	-10.35677694

表 4.6 300 筆訓練資料下之 AIC 值與計算過程

TRV_300	V	lnV	d	C*d/N	AIC
1	0.12771641	-2.057943017	12	0.08	-1.977943017
2	0.077824557	-2.553298256	24	0.16	-2.393298256
3	0.000413393	-7.791110707	36	0.24	-7.551110707
4	1.29785E-05	-11.25222027	48	0.32	-10.93222027
5	1.17051E-05	-11.35548421	60	0.4	-10.95548421
6	1.18831E-05	-11.34039081	72	0.48	-10.86039081
7	7.20712E-06	-11.84044113	84	0.56	-11.28044113
8	7.82288E-06	-11.75845778	96	0.64	-11.11845778
9	8.2438E-06	-11.70604916	108	0.72	-10.98604916
10	3.43739E-06	-12.5807981	120	0.8	-11.7807981
11	6.44453E-06	-11.95227885	132	0.88	-11.07227885
12	5.06294E-06	-12.19356322	144	0.96	-11.23356322
13	7.19227E-06	-11.84250372	156	1.04	-10.80250372
14	5.26042E-06	-12.15529969	168	1.12	-11.03529969
15	4.41021E-06	-12.33158825	180	1.2	-11.13158825
16	4.92482E-06	-12.22122283	192	1.28	-10.94122283
17	5.17839E-06	-12.17101636	204	1.36	-10.81101636
18	4.12355E-06	-12.39879612	216	1.44	-10.95879612
19	4.84276E-06	-12.23802575	228	1.52	-10.71802575
20	3.57843E-06	-12.5405864	240	1.6	-10.9405864

表 4.7 360 筆訓練資料下之 AIC 值與計算過程

TRV_360	V	lnV	d	C*d/N	AIC
1	0.180719095	-1.710811414	12	0.066667	-1.644144747
2	0.127823796	-2.057102556	24	0.133333	-1.923769222
3	0.006005369	-5.11510145	36	0.2	-4.91510145
4	1.83271E-05	-10.90713026	48	0.266667	-10.6404636
5	2.04411E-05	-10.79796347	60	0.333333	-10.46463013
6	1.76293E-05	-10.94594827	72	0.4	-10.54594827
7	1.3199E-05	-11.23536797	84	0.466667	-10.76870131
8	1.05709E-05	-11.45740562	96	0.533333	-10.92407228
9	1.03564E-05	-11.47790684	108	0.6	-10.87790684
10	8.30812E-06	-11.69827721	120	0.666667	-11.03161054
11	7.85882E-06	-11.75387409	132	0.733333	-11.02054076
12	8.22462E-06	-11.70837846	144	0.8	-10.90837846
13	7.74326E-06	-11.76868777	156	0.866667	-10.9020211
14	7.2049E-06	-11.84074921	168	0.933333	-10.90741587
15	9.71813E-06	-11.54151734	180	1	-10.54151734
16	8.64276E-06	-11.65878858	192	1.066667	-10.59212192
17	9.15174E-06	-11.60156653	204	1.133333	-10.4682332
18	8.07175E-06	-11.72714025	216	1.2	-10.52714025
19	7.86292E-06	-11.75335252	228	1.266667	-10.48668585
20	6.71631E-06	-11.91097166	240	1.333333	-10.57763833

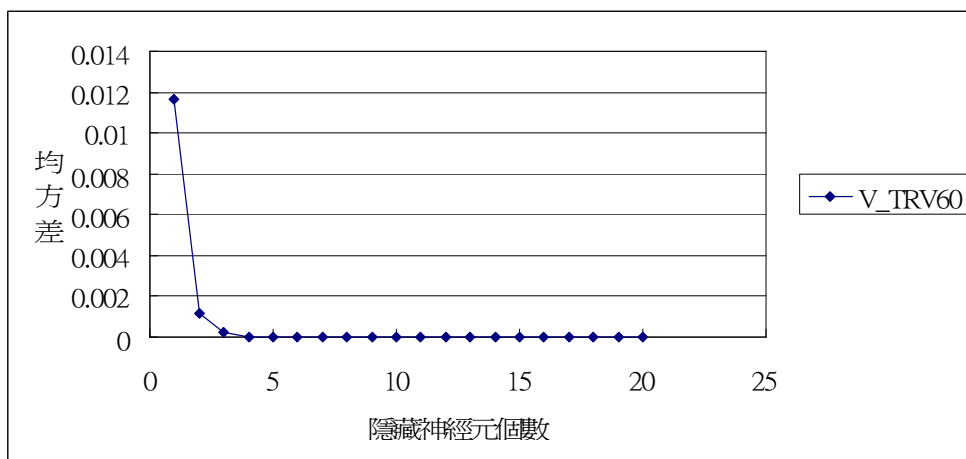


圖 4.2 60 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖

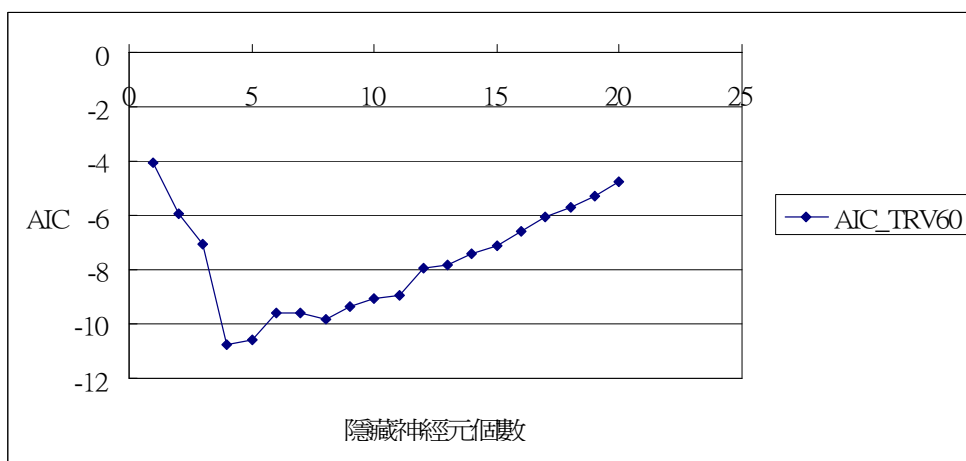


圖 4.3 60 筆訓練資料下隱藏神經元個數與 AIC 之關係圖

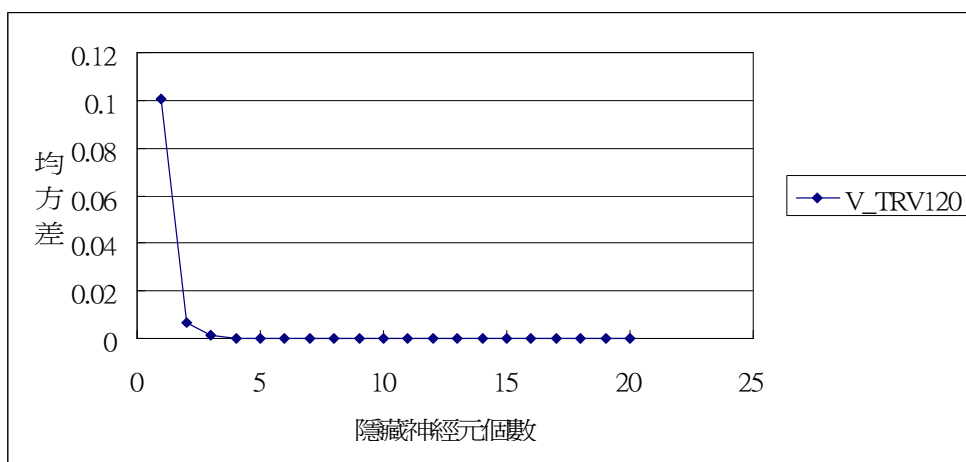


圖 4.4 120 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖

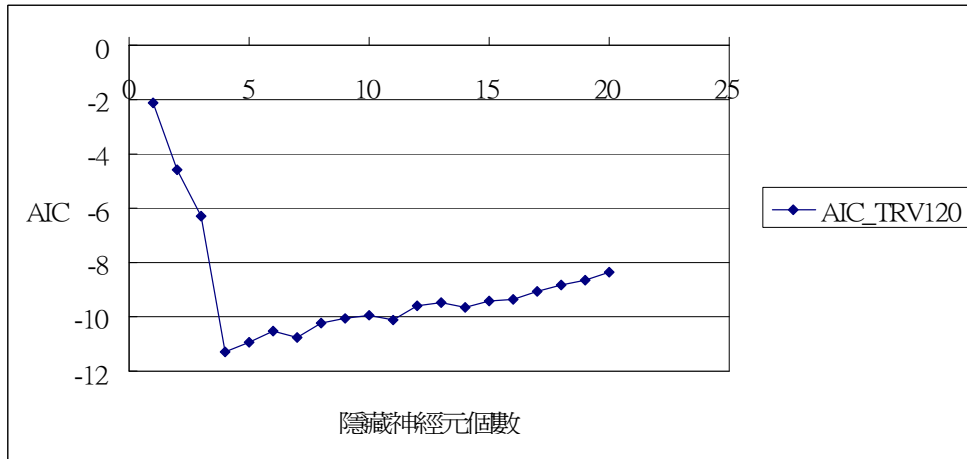


圖 4.5 120 筆訓練資料下隱藏神經元個數與 AIC 之關係圖

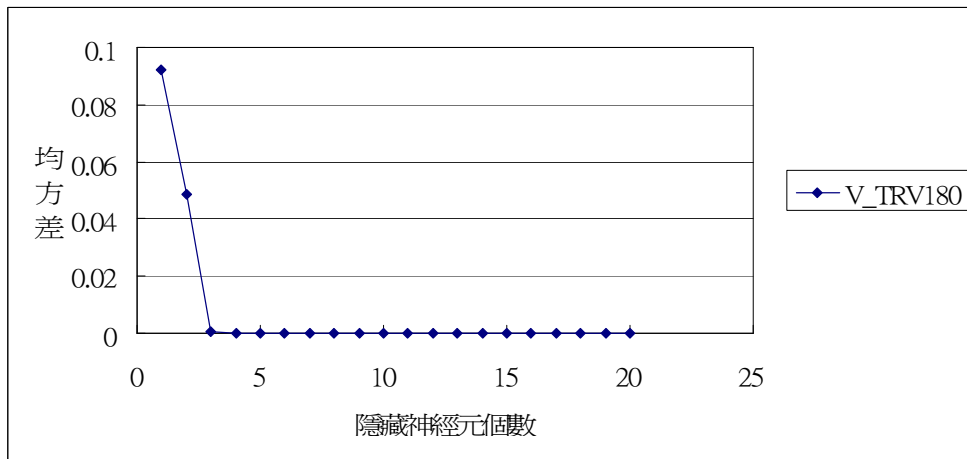


圖 4.6 180 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖

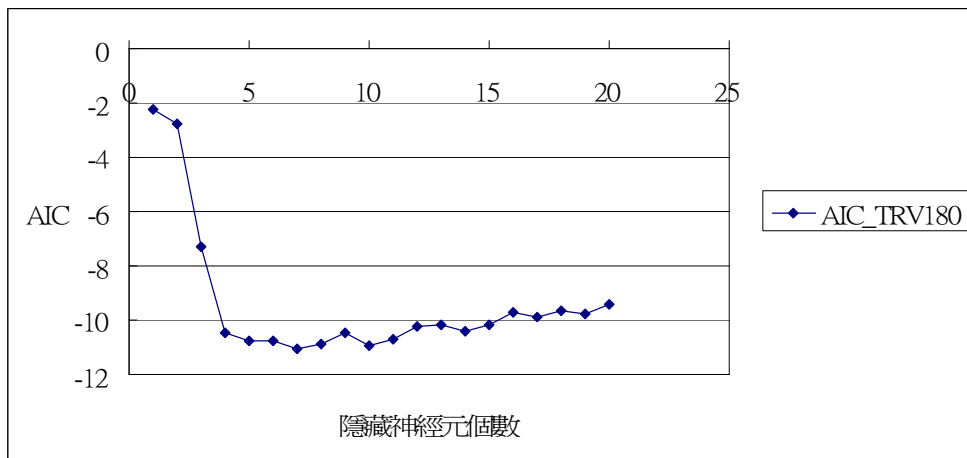


圖 4.7 180 筆訓練資料下隱藏神經元個數與 AIC 之關係圖

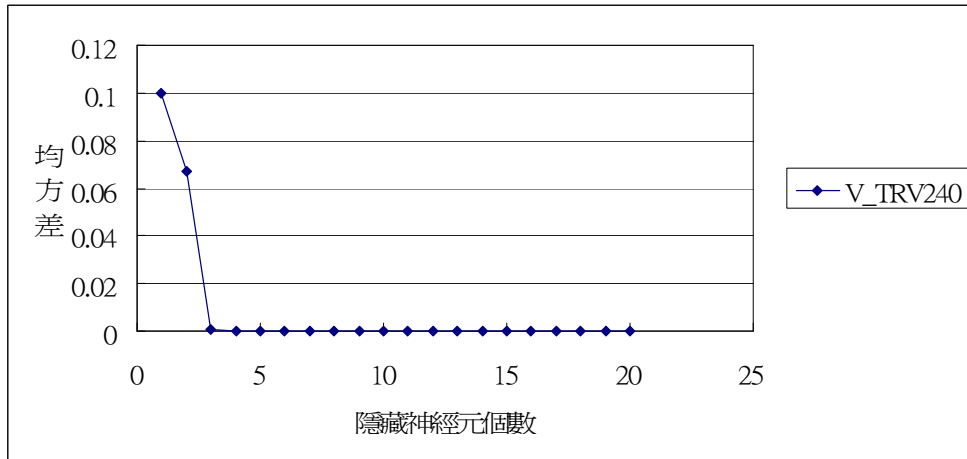


圖 4.8 240 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖

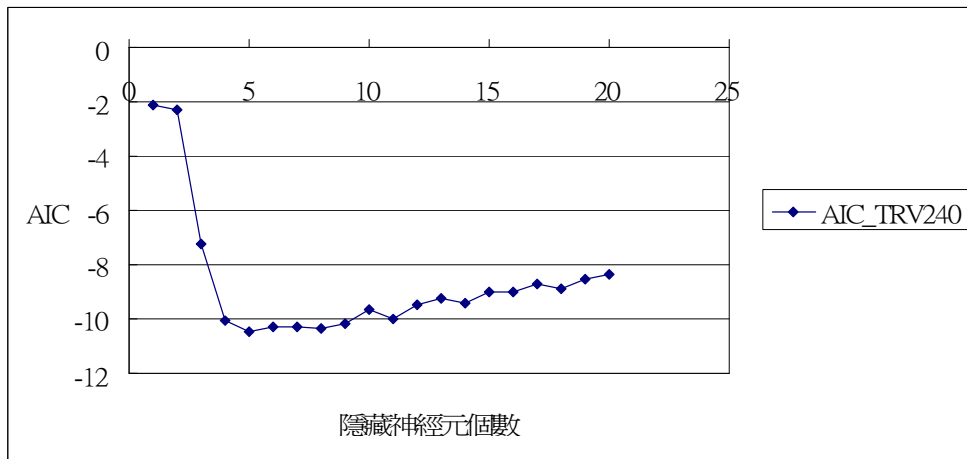


圖 4.9 240 筆訓練資料下隱藏神經元個數與 AIC 之關係圖

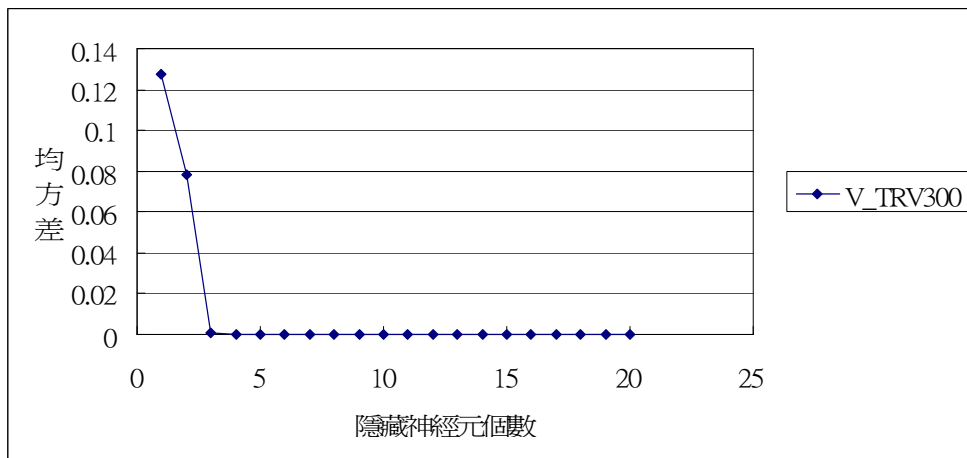


圖 4.10 300 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖

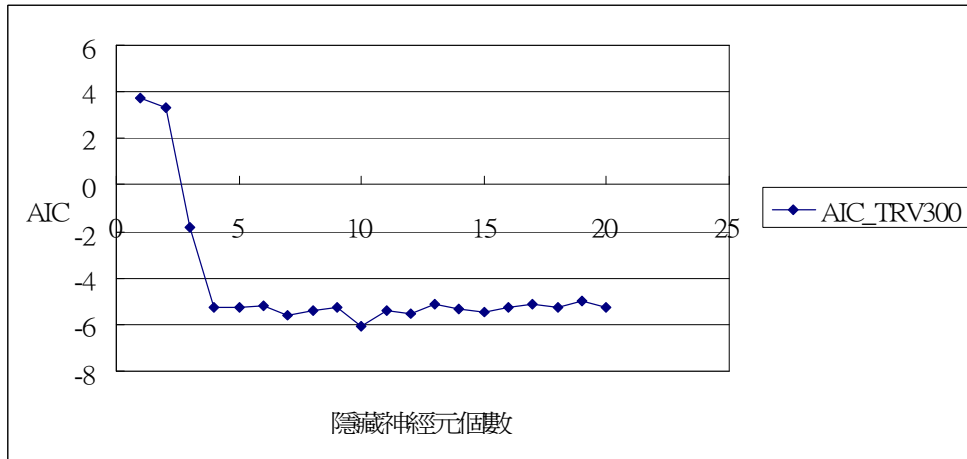


圖 4.11 300 筆訓練資料下隱藏神經元個數與 AIC 之關係圖

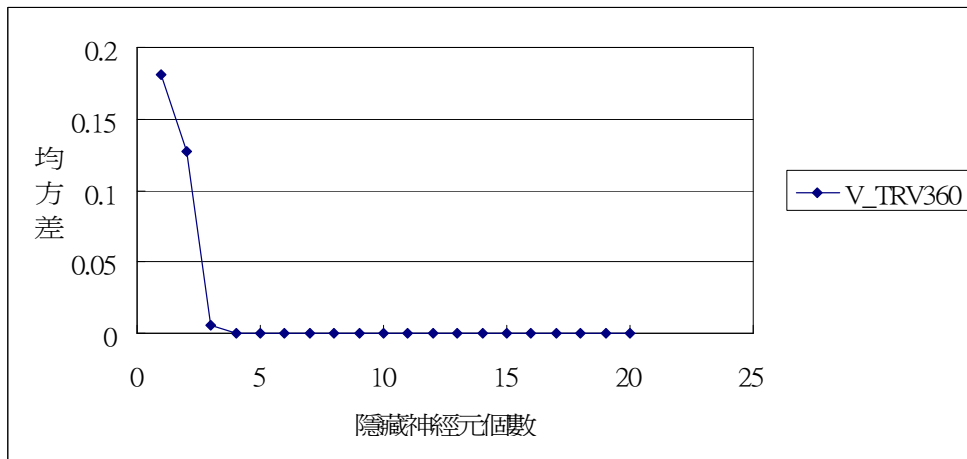


圖 4.12 360 筆訓練資料下隱藏神經元個數與停止誤差值之關係圖

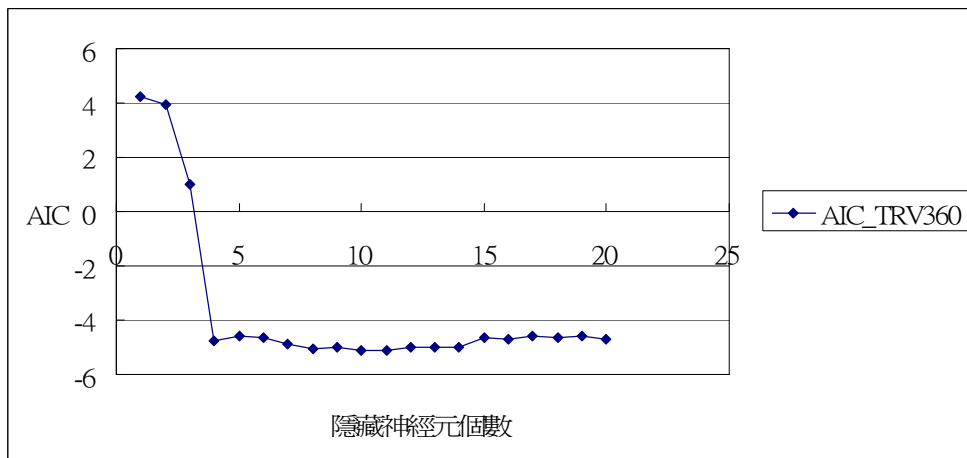


圖 4.13 360 筆訓練資料下隱藏神經元個數與 AIC 之關係圖

4.2.2 小波濾除冗餘權重步驟

1. 首先以 MODFLOW 產生 360 筆資料，並且將之正規化，以供類神經網路訓練之用，每筆資料包含 8 個輸入向量與 4 個輸出輸出向量。
2. 以 MATLAB 程式設計類神經網路，使用最普通最平常的倒傳遞網路，各參數的設定如下所示：
 - (1) 隱藏層一層，活化函數使用 tansig，為雙彎曲正切函數；
 - (2) 輸出層活化函數使用 purelin，為線性函數；
 - (3) 訓練法使用 trainrp，稱為 Resilient Backpropagation 學習法；
 - (4) 學習速率設為 0.001，動量係數為 0.9；
 - (5) 訓練次數為 50000 次，訓練目標精度 $MSE=0.00001$ ；
 - (6) 初始權重與初始偏權值皆為 random。
3. 取出事先準備好的資料的前 120 筆，使用上述的類神經網路開始訓練，隱藏神經元個數設為 20 個（極大）開始訓練；因為倒傳遞類神經網路之結果為多值解，隨機的初始權重亦可能影響最後停止誤差之大小，為了能夠得到各個不同隱藏層神經元個數下的最佳停止誤差值與最佳權重值，所以每個神經元個數之訓練皆循環 5 次，取停止誤差值最小的該次權重為最佳權重進行小波分析。
4. 使用 MATLAB 程式套裝小波程式進行分析，使用 dpl（離散小波轉換），小波函數使用 db3（如圖 4.14 所示），相關細節說明如後：
 - (1) 欲將權重輸入小波裡過濾必須先將之排序，本論文採取輸入層權重在前，輸出層權重在後的排法，先將所有的輸入層權重排完，再排輸出層權重，共計有四種排法（參考圖 4.15～圖 4.18）；因為以小波過濾權重以決定隱藏神經元個數之作法本研究實屬先例，故本論文僅先取第二種（如圖 4.16 所示）嘗試研究。
 - (2) 在眾多的小波裡選用 db3 之原因在於其圖形有個非常明顯的突起，因此當其在待分析之序列上平移之時，經過摺積之計算將會使得離零值越多之值越

大，凸顯訊號中之特徵值，抑制過小之雜訊；亦即不論正負，取絕對值大者保留，利用此一性質刪去類神經網路裡權重值過小之連結。唯要注意的是雖然圖形非常相像，但 db3 是不對稱的，而墨西哥草帽函數是對稱的，兩者並不相同，本論文選用的是 MATLAB 套裝程式的 db3。

(3)本研究在訓練類神經網路時，初始權重皆為-1~1 之間的 Random 值，故推想最終訓練結束後之權重，其相互間之差值之絕對值應該不會有過大之變化，因此最初小波過濾門檻值設為 0.5，往上往下各增減 0.1，以嘗試各種不同門檻值；發現若在 0.7 以上則權重會盡數被打斷，若在 0.5 以下則權重濾除之效果不明顯，因此最後決定門檻值為 0.6。

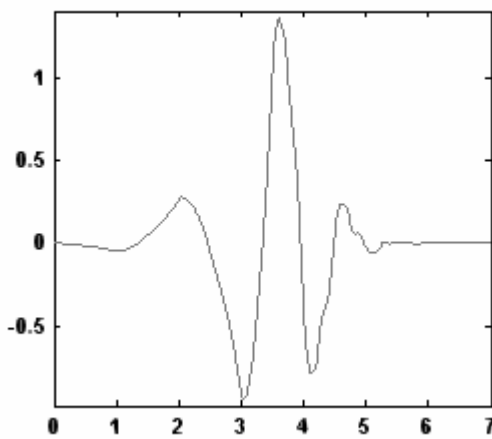


圖 4.14 (a) db3 函數圖

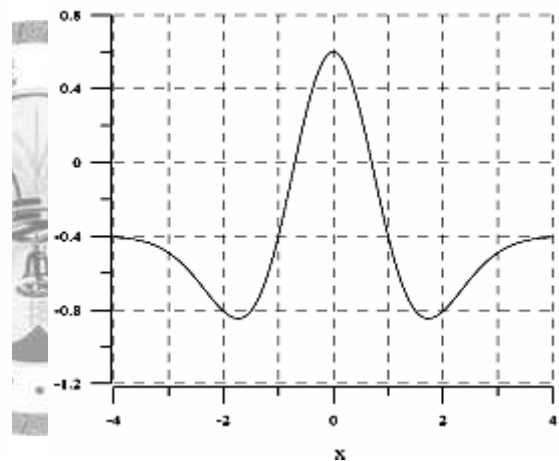
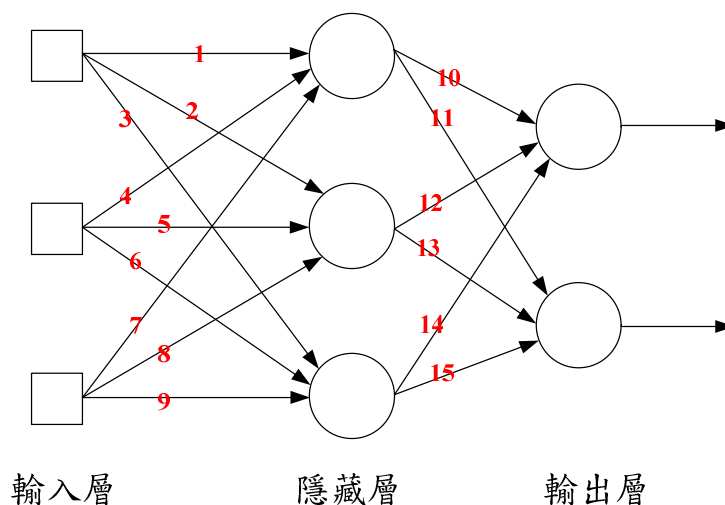


圖 4.14 (b) 墨西哥小帽函數圖



輸入層 隱藏層 輸出層

圖 4.15 權重排列方式一

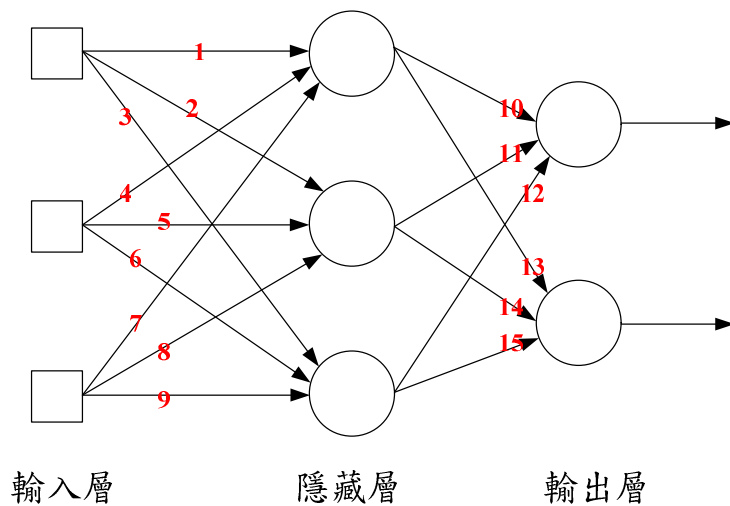


圖 4.16 權重排列方式二

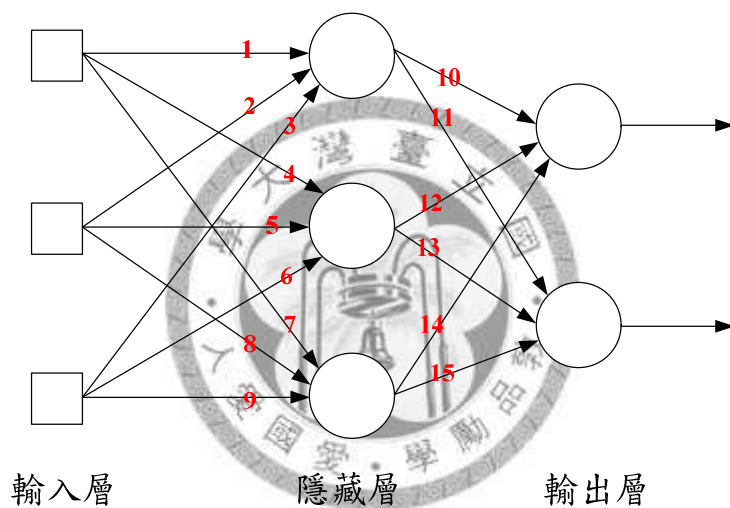


圖 4.17 權重排列方式三

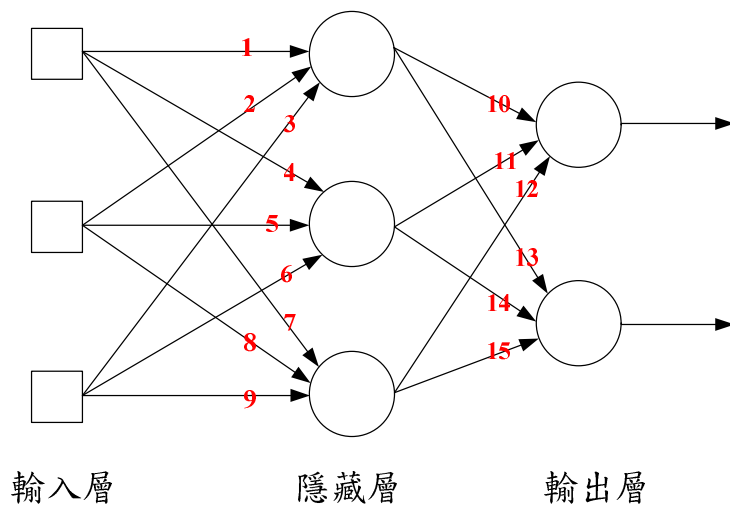


圖 4.18 權重排列方式四

5.將前一步驟所得到的 120 筆資料下,20 個隱藏神經元下的最佳權重值放入小波程式裡過濾,被濾除的權重處就補零;只要發現有某個隱藏神經元的連結權重全部被打斷,就視這個隱藏神經元為不必要的隱藏神經元,相關說明如下:

(1)只有權重盡數被打斷的隱藏層神經元會被移除,換言之,即便只剩下一個連結權重沒被打斷,該隱藏神經元仍然會被保留(相關之示意圖可以參考圖 4.19 與圖 4.20 所示);如此,每次濾除之神經元個數不再限制為一個,可同時濾除多個不必要的隱藏層神經元。

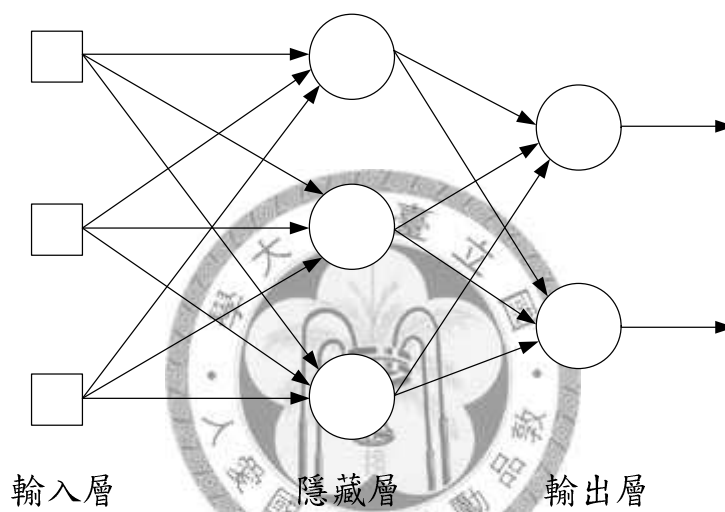


圖 4.19 權重濾除前

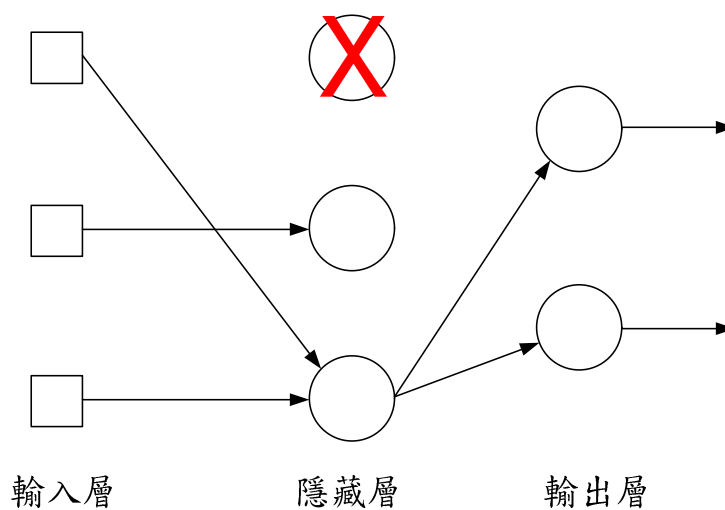


圖 4.20 權重濾除後

表 4.8 小波濾除權重範例

II	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	<1>	<2>	<3>	<4>
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0.44504	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	-0.61357	0	0	0	0	0	0	0	0	-0.39749
10	0	0	0	0	0	0	0	0	1.54588	0	0	0
11	0	0	0	0	0	0	0	0	-0.31612	0	0	0.91093
12	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	-1.09613
14	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0.85401
16	0	0	0.71741	0	0	0	0	0	0	0	0	0
17	0	0	-1.1934	0	0	0	0	0	0.73518	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0

(2)參考上頁表 4.8，此為 120 筆訓練資料下，20 個隱藏神經元時，將權重以小波過濾後的結果；表上方之(1)~(8)表示輸入層神經元，<1>~<4>表示輸出層神經元，表左方之 1~20 表示隱藏層神經元，表左上方之 II 表示在 5 次的訓練停止誤差裡，第二次是最小的；而本表當中有 12 列完全為零，故我們可以說在 120 筆訓練資料 20 個隱藏神經元的情形下，有 12 個隱藏神經元是不必要的，可將隱藏神經元數減少為 8 個。

6.如此依序減少隱藏神經元個數，直到某一隱藏神經元個數下，再沒有隱藏神經元的權重全部被打斷的情況下即停止，而此時的隱藏神經元個數即視為 120 筆資料下的最佳隱藏神經元個數。

7.重複步驟 3.~5.便可得到資料筆數為 120、180、240、300、360 筆情況下各自的最佳隱藏神經元個數，結果如表 4.9 與圖 4.21 所示。

資料筆數	門檻值	隱藏神經元個數決定過程				
		20	8	5		
120	0.6	20	8	5		
180		20	9	7	6	5
240		20	7			
300		20	7			
360		20	10			

表 4.9 不同資料筆數下隱藏神經元個數決定過程

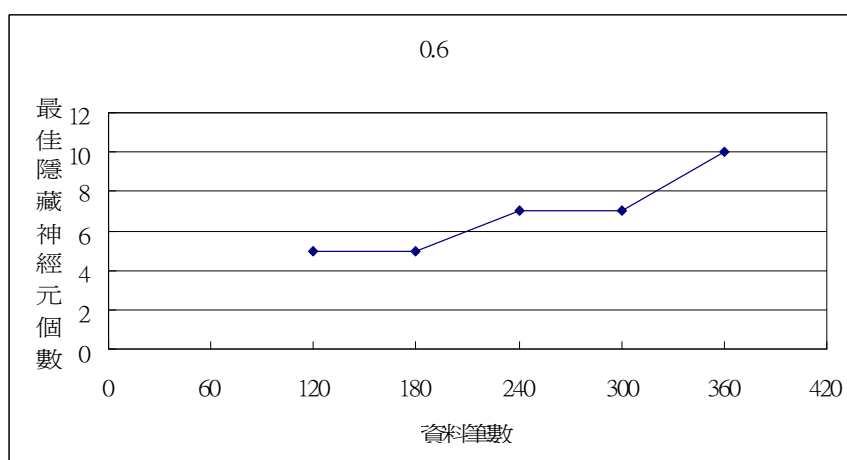


圖 4.21 不同資料筆數下最佳隱藏神經元個數關係圖（小波方法）

4.3 AIC 指標方法實驗結果討論

1.由圖 4.2、圖 4.4、圖 4.6、圖 4.8、圖 4.10、圖 4.12 可得知，以本範例而言，類神經的訓練是成功的，因為圖中所呈現的趨勢與一般普遍認知的類神經網路之訓練結果相吻合：當隱藏神經元個數逐漸增加，訓練停止時的均方差也隨之減小，最後收斂於某定值附近，不再隨隱藏神經元個數增加而有明顯減少。

2.由圖 4.3、圖 4.5、圖 4.7、圖 4.9、圖 4.11、圖 4.13 分析，發現 AIC 指標與類神經網路結合應用於本範例的結果普通，因為從圖 4.7，也就是訓練資料筆數為 180 筆開始，很明顯地曲線並沒有辦法如同理論曲線所預測地往上升；針對此一現象逐一分析如後。

(1)從 AIC 的公式可以看出，前項是均方差相關項，後項是參數維度與資料筆數項，如果後項的增加值無法大過均方差項的減少值，曲線當然就無法有明顯上升的趨勢。

(2)在本例中，從訓練資料筆數為 180 筆開始，造成參數維度項增值不夠大的主要原因在於訓練資料筆數過多；因此若將訓練資料筆數為 180 筆時之 C 值調為 3，而訓練資料筆數為 240 筆時之 C 值調為 4，那麼 AIC 曲線就能有較明顯地上升，如圖 4.22 與圖 4.23 所示；唯其上升趨勢仍不及訓練資料筆數為 60 筆與 120 筆時來得顯著。

(3)是否應該將訓練資料筆數為 300 筆與 360 筆之 C 值予以升高？根據傳統的經驗，C 值以 2~4 為佳，故對於訓練資料筆數為 300 筆與 360 筆之 C 值，本論文覺得不應予以升高，反而應考慮是否將本例中訓練用的資料筆數上限定為 240，以防止模式的過度參數化。

3.傳統上應用 AIC 指標時，習慣以 AIC 值最小的模式為最佳模式，故如果不考慮曲線無法上升的情況，直接以 AIC 值為最小的隱藏神經元數為最佳者，可得如表 4.10 與圖 4.24 之結果。此一結果尚算正確，因為它顯示出當訓練資料個數增加時，模式需要更多隱藏神經元個數（意即：參數）方能描述問題。

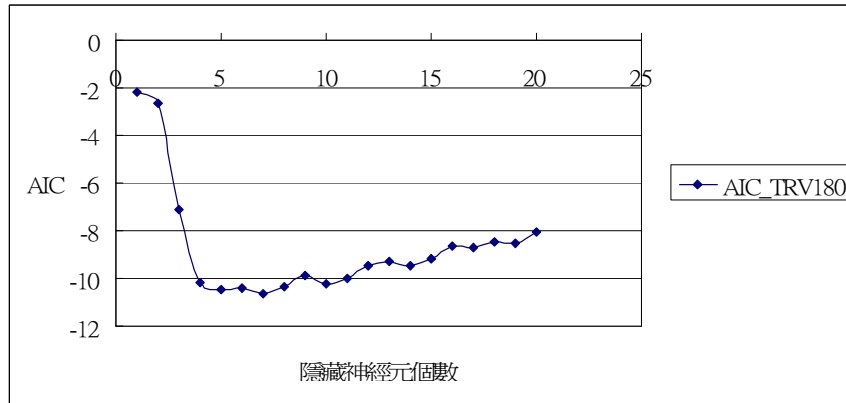


圖 4.22 180 筆訓練資料下隱藏神經元個數與 AIC 之關係圖 (C=3)

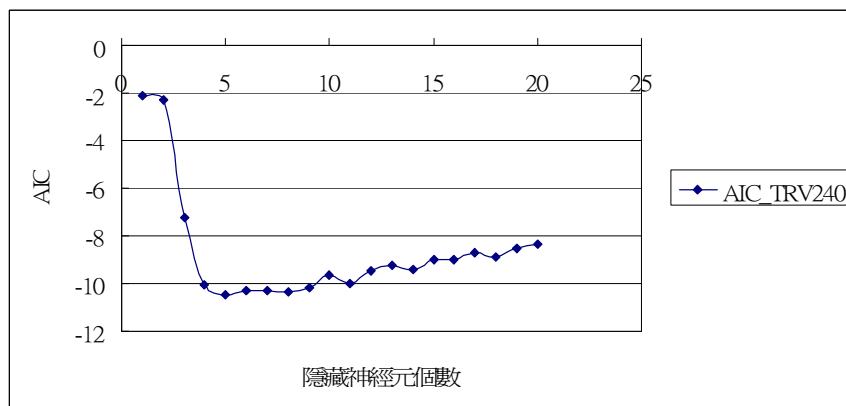


圖 4.23 240 筆訓練資料下隱藏神經元個數與 AIC 之關係圖 (C=4)

訓練資料筆數	60	120	180	240	300	360
最佳隱藏神經元個數	4	4	7	8	10	10

表 4.10 訓練資料筆數與最佳隱藏神經元數 (AIC 方法)

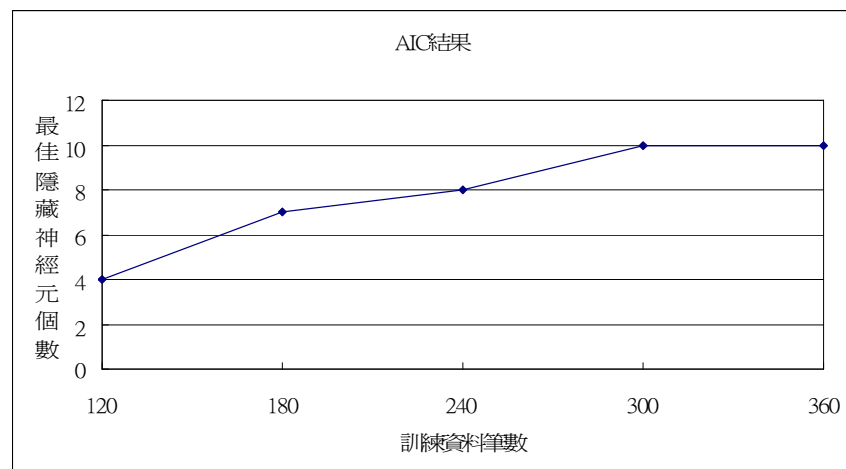


圖 4.24 訓練資料筆數與最佳隱藏神經元數關係圖 (AIC 方法)

4.4 小波濾除冗餘權重結果討論

- 1.根據第 52 頁表 4.9 與圖 4.21 所示，隨著訓練資料筆數的增加，模式所需要的隱藏神經元個數亦增加；此與一般的認知相符，表示當訓練資料個數增加時，模式需要更多隱藏神經元個數（意即：參數）方能描述問題。
- 2.以上討論之初始隱藏神經元個數皆為 20，於是本研究想探討，若是從其他的初始隱藏神經元個數開始刪減，其最後結果又如何？故實驗之，得結果如表 4.11～表 4.15 所示，發現即便在相同的訓練資料筆數下，最後之最佳隱藏神經元個數並不統一；但如果取停止次數最多的神經元數為最佳隱藏神經元個數，則可整理得表 4.16 並作圖如圖 4.25 所示，仍顯示隨著訓練資料筆數的增加，模式所需要的隱藏神經元個數亦增加，此結果與一般的認知相符；因此即使最後之最佳隱藏神經元個數並不統一，本論文之方法論仍然無誤，而且可行性是高的。

表 4.11 120 筆訓練資料下權重過濾後神經元減少路徑

資料數	門檻值	隱藏神經元個數決定過程					
		20	8	5			
120	0.6	19	12	7			
		18	7				
		17	6	5			
		16	5				
		15	10	8	5		
		14	11	10	8	5	
		13	11	10	8	5	
		12	7				
		11	10	8	5		
		10	8	5			
		9	7				
		8	5				
		7	7				
		6	5				
		5	5				

表 4.12 180 筆訓練資料下權重過濾後神經元減少路徑

資料數	門檻值	隱藏神經元個數決定過程					
180	0.6	20	9	7	6	5	
		19	13	7	6	5	
		18	6	5			
		17	10	9	7	6	5
		16	11	5			
		15	6	5			
		14	10	9	7	6	5
		13	7	6	5		
		12	10	9	7	6	5
		11	5				
		10	9	7	6	5	
		9	7	6	5		
		8	7	6	5		
		7	6	5			
		6	5				
		5	5				

表 4.13 240 筆訓練資料下權重過濾後神經元減少路徑

資料數	門檻值	隱藏神經元個數決定過程					
240	0.6	20	7				
		19	10	9	7		
		18	13	7			
		17	7				
		16	12	10	9	7	
		15	12	10	9	7	
		14	8				
		13	7				
		12	10	9	7		
		11	7				
		10	9	7			
		9	7				
		8	8				
		7	7				
		6	6				
		5	5				

表 4.14 300 筆訓練資料下權重過濾後神經元減少路徑

資料數	門檻值	隱藏神經元個數決定過程					
300	0.6	20	7				
		19	13	10	7		
		18	11	9			
		17	13	10	7		
		16	9				
		15	12	8			
		14	7				
		13	10	7			
		12	8				
		11	9				
		10	7				
		9	9				
		8	8				
		7	7				
		6	6				
5	5						

表 4.15 360 筆訓練資料下權重過濾後神經元減少路徑

資料數	門檻值	隱藏神經元個數決定過程					
360	0.6	20	10				
		19	9	8			
		18	9	8			
		17	17				
		16	7				
		15	9	8			
		14	12	9	8		
		13	10				
		12	9	8			
		11	9	8			
		10	10				
		9	8				
		8	8				
		7	7				
		6	6				
5	5						

門檻值=0.6		資料筆數				
		120	180	240	300	360
停止時之 隱藏神經 元個數	5	10	16			
	7	5		14	7	2
	8			2	3	8
	9				4	
	10					3
	17					1

表 4.16 訓練資料筆數與最佳隱藏神經元數（小波方法，門檻值=0.6）

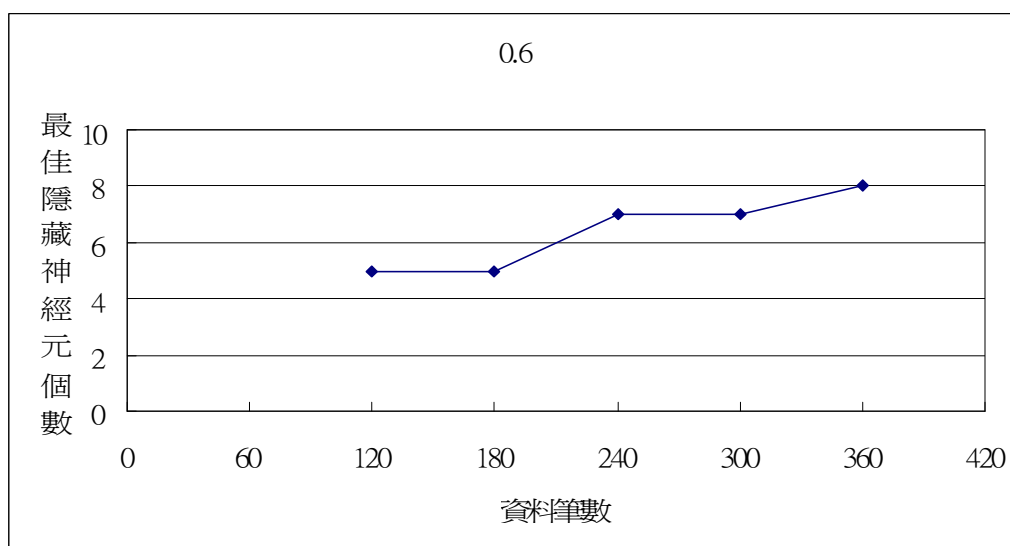


圖 4.25 訓練資料筆數與最佳隱藏神經元數關係圖（小波方法，門檻值=0.6）

- 3.各訓練資料筆數下的隱藏神經元個數無法各自統一在某一定值此一現象，可能的原因在於，類神經網路訓練之初，初始權重值與初始偏權值之猜值皆為隨機給定，故造成訓練完成之後，權重值之大小與分佈帶有隨機的特性，導致使用一固定之小波門檻值去過濾時，會有振盪結果發生。
- 4.是否應該配合權重值大小變動的隨機特性，亦採用大小變動的小波門檻值？本論文認為不妥；因為一旦採用變動式的門檻值，將使得過濾的過程無一統一標準，再者誠如前面章節所述，初始權重猜值雖為隨機給定，但其值皆被限制在-1~1之間，因此類神經網路訓練完成之後其相互間之差值應該也有限，故不

需特地再以變動式之門檻值予以過濾與分析。

5. 重新檢視表 4.11~表 4.15，發現不論訓練資料筆數為多少，隱藏神經元個數為 5 之時，都沒有隱藏神經元的連結權重盡數被打斷，此一現象可說明為了能夠描述本實驗的資料，網路至少需要 4 個以上之隱藏神經元，方可獲致一較為穩定的精度。
6. 重新整理 4.11~表 4.15 得出下表 4.17，可發現不論訓練資料筆數為多少，初始隱藏神經元個數越大，被小波濾除的隱藏神經元個數普遍較多，這顯示隨著隱藏神經元個數的增加，網路中之權重連結有變小並分散的趨勢，再次驗證本研究將過小權重剔除以減少並優選隱藏神經元個數之作法，其過程合理而且結果可信；唯受限於權重初始猜值為亂數之影響，被濾除的隱藏神經元個數有振盪情形發生，無法完全與隱藏神經元個數的多寡成正比。

門檻值 0.6						
被小波濾除的 神經元個數		訓練資料筆數				
		120	180	240	300	360
隱藏神經元個數	20	12	11	13	13	10
	19	7	6	9	6	10
	18	11	12	5	7	9
	17	11	7	10	4	0
	16	11	5	4	7	9
	15	5	9	3	3	6
	14	3	4	6	7	2
	13	2	6	6	3	3
	12	5	2	2	4	3
	11	1	6	4	2	2
	10	2	1	1	3	0
	9	2	2	2	0	1
	8	3	1	0	0	0
	7	0	1	0	0	0
	6	1	1	0	0	0
5	0	0	0	0	0	

表 4.17 不論訓練資料筆數被濾除之隱藏神經元個數 (門檻值=0.6)

4.5 AIC 指標方法與小波濾除冗餘權重方法之綜合比較

1.將 AIC 指標方法與小波濾除冗餘權重方法之結果整理，得表 4.18 與圖 4.26；

觀察後可發現，除了在訓練資料筆數 120 組之時，AIC 指標方法的最佳隱藏神經元個數 4 個少於小波分析方法的 5 個之外，剩於的訓練資料筆數情況下，AIC 指標法方所優選出來的最佳隱藏神經元個數皆多於小波分析方法所優選出來的個數。

		訓練資料筆數				
		120	180	240	300	360
最佳隱藏神經元個數	AIC 指標方法	4	7	8	10	10
	小波分析方法	5	5	7	7	8

表 4.18 AIC 與小波方法結果比較

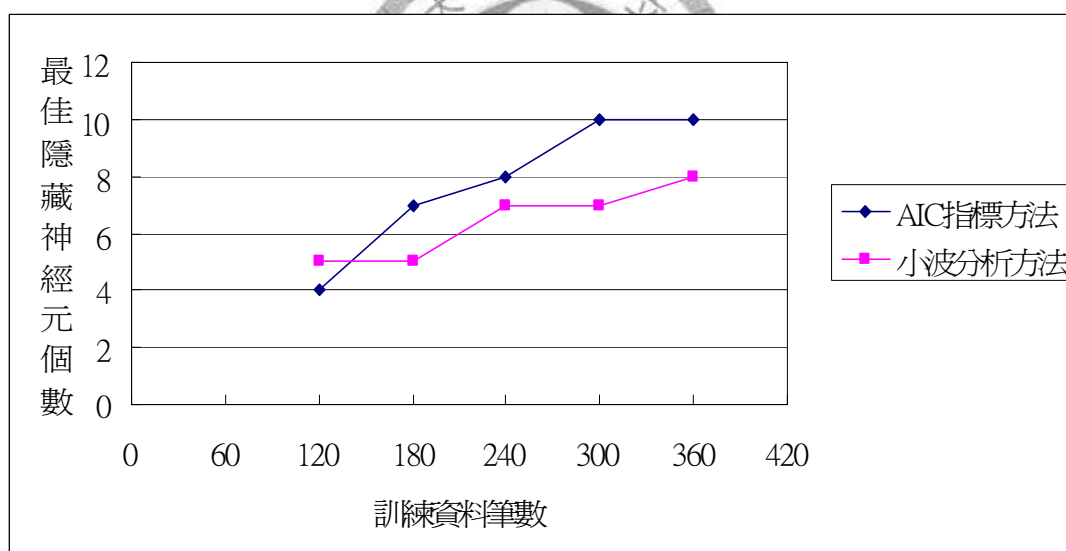


圖 4.26 AIC 與小波方法結果比較

2.針對兩種實驗方法所決定出來的最佳隱藏神經元數不一此種情形，以下提出可能的原因與分析。

(1)不同的指標在優選模式之參數時各自有其不同的考量因素，因此若最後所有指標的結果能收斂在同一個模式當然是最好，若不然，則可視各自之情形，討論其結果是否合理。

(2)仔細觀察圖 4.2、圖 4.4、圖 4.6、圖 4.8、圖 4.10、圖 4.12 可發現，網路在隱藏神經元個數達 2 個開始，之後便急速收斂，到了隱藏神經元個數為 4 之後，最後之停止誤差值更是幾乎不會再有大變動，本論文不論是 AIC 指標，或是小波濾除方法，其最後所選出來的隱藏神經元個數皆大於 4，足以顯示本研究之實驗結果是合理的。

3.類神經網路之初始權重猜值大多隨機給定，因此最後之結果不論是對停止誤差而言，或是對最佳之隱藏神經元個數而言，勢必都會有部份的小震盪；但因為最終網路之誤差皆能達到吾人所要求之精度，所以即使本論文之方法論與傳統的 AIC 指標方法所優選出來之隱藏神經元個數不同，本研究所提出之方法論於實用上仍相當可行；而且使用小波快速濾除冗餘神經元以決定最佳隱神經元個數，其時間上之經濟效益是傳統窮舉法所無法比擬的。

4.以上所討論的範例，輸入小波模式過濾的權重排列方式皆採用圖 4.15 的排列方式，過濾之門檻值取 0.6，因為 0.5 與 0.7 之實驗結果皆令人不滿意（參考圖 4.27 所示）；本論文曾經嘗試分析圖 4.15、4.17、4.18 等三種權重排列方式，但是不論門檻值為 0.5、0.6、或 0.7，其結果皆不如圖 4.16 的排列方式令人滿意；未來或許可以嘗試分析更多不同門檻值之實驗結果，進而找出 AIC 指標與小波門檻值之間關聯性。

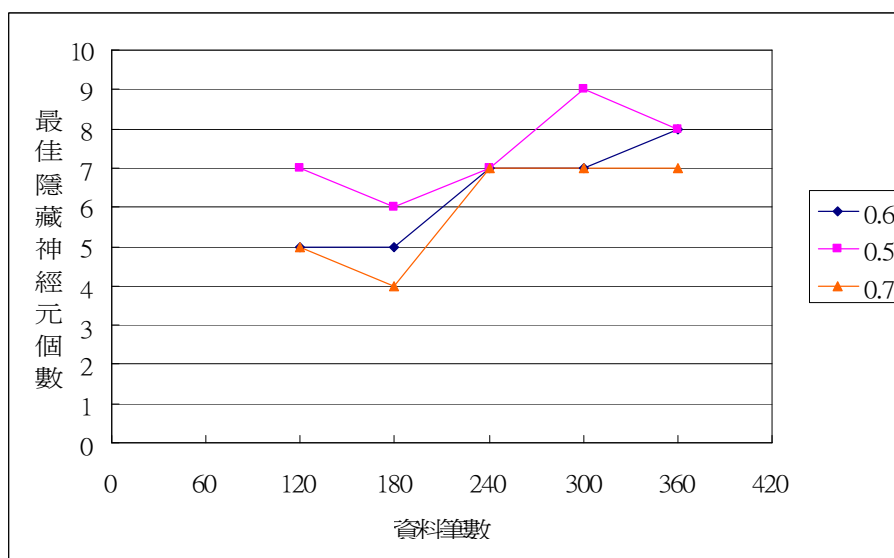


圖 4.27 小波方法不同門檻值對最佳隱藏神經元個數之影響

5.類神經網路之初始權重值為隨機給定此點，在本研究的過程與結果分析上帶來極大的不便，因此亦曾嘗試固定初始權重值，然此一舉動對實驗結果與後續分析並無幫助，原因如下：

(1)參考附錄一，經過仔細觀察可以發現，所有的權重修正項皆是初始權重之函數，因此如果將同一隱藏神經元個數下的初始權重值與初始偏權值皆固定，那麼不管訓練循環幾次，最後的結果都將相同。

(2)又在不同的隱藏層神經元個數下，因為隱藏層神經元個數不同，那整體函數的運作也會有很大的差異，因此對於倒傳遞類神經網路而言，對不同的隱藏神經元個數是無法有一統一初始值的；再者，仔細觀察附錄一之結果可發現，一般的多項式在項數不同時或許可統一以原點為初始值；但是在倒傳遞類神經網路中，一旦所有的權重初始值皆給為零，將始得網路節點的輸出值 ($f(net)$) 與權重修正項 (Δw) 變為零，致使類神經網路之運作呈現空轉狀態，即便訓練次數增加為幾千幾萬次，停止誤差值都不會改善。

6.從第 39 頁的表 4.1 可發現，即便每個隱藏神經元皆重覆訓練五次以求最小之均方差，但結果發現最後的停止均方差仍然呈現振盪狀態，並未隨著隱藏神經元個數之增加而持續減少，故未來或許只訓練一次即可，不需要重覆五次。

7.類神經網路受到廣泛應用之原因在於，它可以在不清楚欲分析對象之物理意義的情形下，僅靠觀測之輸入與輸出成對資料 (input and output data sets) 去建構一高精度的模擬模式，因此其運作上可以說是一種統計模式；再者，由附錄一可知類神經網路最後訓練完畢之連結權重值與偏權值，其與初始值有很大的關連性；因此本研究認為，類神經網路之運作與網路中之連結權重值與偏權值，其物理意義是相當薄弱的。唯一可知者是，在本研究的分析過程中發現隨著隱藏神經元個數之增加，網路之連結權重有分散並且變小的現象；若欲更加詳細地分析與確認倒傳遞類神經網路之運作與各連結權重值與偏權值是否有物理意義，或許可待後人研究，本論文不再深入探討。

第五章 結論與建議

5.1 研究結論

本研究將倒傳遞類神經網路之隱藏神經元個數的決定，視為一般參數檢定的過程，使用 AIC 指標來決定隱藏神經元個數並優選模式。同時，因為類神經網路的各個權重 w_i 主要是模擬不同生物神經元間的連結強弱，權重為正且越大表示連結越強且增益越大；權重為負且越小表示連結雖強但是為抑制的功用；若權重接近零，表示連結很弱。因此本研究分析在不同隱藏層神經元個數下，各個權重值的大小之變化趨勢，將每一組權重視為一系列的頻譜，找出一適合的門檻值，使用小波分析將之過濾，把過於接近零，也就是連結度越弱的權重予以剔除，進而得出一種更加切確且快速的方式以決定隱藏層神經元個數，以期建立一精確之類神經模式以模擬地下水水位；相關實驗結果與結論整理如後：

1. 本論文經過數值實驗之驗證，證明第三章所提出之方法論為有效，同時在本研究之實驗範例中，建議小波過濾門檻值設為 0.6 最為合適，輸入小波模式裡分析之權重序列採用圖 4.16 之排列方式。
2. 初始隱藏神經元個數越多，被小波濾除的隱藏神經元個數普遍較多，這顯示隨著隱藏神經元個數的增加，網路中之權重連結有變小並分散的趨勢；唯受限於權重初始猜值為亂數之影響，被濾除的隱藏神經元個數有振盪情形發生，無法完全與隱藏神經元個數的多寡成正比（可參考表 4.17）。
3. 類神經網路之初始權重猜值大多隨機給定，因此最後之結果不論是對停止誤差而言，或是對最佳之隱藏神經元個數而言，勢必都會有部份的小震盪；但因為最終網路之誤差皆能達到吾人所要求之精度，所以即使本論文之方法論與傳統的 AIC 指標方法所優選出來之隱藏神經元個數不同，本研究所提出之方法論

於實用上仍相當可行；而且使用小波快速濾除冗餘神經元以決定最佳隱神經元個數，其時間上之經濟效益是傳統窮舉法所無法比擬的。

5.2 建議事項

1. 因為本研究以小波過濾類神經網路權重之作法乃屬先例，故輸入小波模式之權重序列可能的四種排列方式皆有探討，但是不夠深入；未來如果對本論文之方法有興趣者，可以進一步深入探討；例如嘗試 db1 或 db2 等不同的小波函數，抑或使用更多不同之門檻值，更甚至可進一步探討偏權值對於本論文之方法論是否有所影響。
2. 本論文所使用之類神經網路為最傳統的倒傳遞類神經網路，根據文獻回顧可知其他種類之類神經網路，應用於「降雨—逕流」等其他水利方面都有不錯之效果，相信若嘗試將這些不同的類神經網路應用於地下水水位之模擬，應該也能獲致不錯之研究成果。
3. 本研究因重點著重於優選隱藏神經元之個數，故輸入類神經網路之訓練資料僅有淨抽水量與前一時刻水位；未來或許可用其他之參考因素為輸入資料以預測地下水水位，例如輸入週邊測站來推估特定觀測站之地下水水位，又如使用高程、相對距離、雨量等資料做為輸入資料，相信皆具有其研究價值。
4. 因為類神經網路訓練時，初始權重猜值皆為隨機給定，故以小波分析最後之權重，其結果難免受到隨機因素之影響而有所振盪；未來或許可以嘗試將小波融入類神經網路架構中，從內部直接改善優化類神經網路，甚至發展出不同於現下任一種架構之類神經網路，都是有可能的。

參考文獻

1. Abdalla Shigidi and Luis A. Garcia, M., 2003. "Parameter estimation in groundwater hydrology using artificial neural networks," *Journal of computing in civil engineering*, ASCE, p.281-289
2. Abrahart R. J., See L. and Kneale P. E., 1998. "New Tools for Neurohydrologists: Using Network Pruning and Model Breeding Algorithms to Discover Optimum Inputs and Architectures," *In Proceedings of the 3rd International Conference on Geocomputation*, University of Bristol.
3. Akaike Hirotugu, 1974. "A new look at the statistical model identification". *IEEE Transactions on Automatic Control*, **19** (6), p.716-723.
4. Asaad, Y. S., M. C. Kieran and G. C. Liang, 1997. "Methods for combining the outputs of different rainfall-runoff models," *Journal of Hydrology*, 197, p.203-229.
5. Bayazit, M. and H. Ahsoy, 2001. "Using wavelets for data generation," *J. Appl. Statist.*, 28(2), 157-166.
6. Cameron, M. Z., H. B. Donald and P. S. Iobodan, 1999. "Short term stream flow forecasting using artificial neural networks," *Journal of Hydrology*, 214, p.32-48.
7. Coulibaly, P., F. Anctil, P. Rasmussen and B. Bobée, 2000. "A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff," *Hydrol. Process*, 14, p.2755-2777.
8. French, M. N., W. F. Krajewski and R. R. Cuykendall, 1992. "Rainfall forecasting in space and time using a neural network," *Journal of Hydrology* 137, p.1-31.
9. Hebbian D. O., 1949. *The Organization of Behavior. A Neuropsychological*. New York: Wiley.

10. Hopfield J. J. and Tank D., 1985. "Neural computations of decisions in optimization problems," *Biological Cybernetics*, 51, p.141-152
11. Hsu, K. L., V. G. Hoshin and S. Soroosh, 1995. "Artificial neural network modeling of the rainfall-runoff process," *Water Resources Research*, 29(10), p.3535-3541.
12. Ioannis N. Daliakopoulos, Paulin Coulibaly, Ioannis K. Tsanis, 2005. "Groundwater level forecasting using artificial neural networks," *Journal of Hydrology*, 309, p.229-240.
13. Kwok T. Y. and Yeung D. Y., 1997. "Constructive Algorithms for Structure Learning in Feedforward Neural Networks for Regression Problems," *IEEE Transactions on Neural Networks*, 3, p.630-645.
14. Labat, D., R. Ababou and A. Mangin, 2000. "Rainfall-runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses," *Journal of Hydrology*, 238, p149-178.
15. Loke, E., E. A. Warnaars, P. Jacobson, F. Nelen and M. A. DoC'eu, 1997. "Artificial neural networks as a tool in urban storm drainage," *Wat. Sci. Tech.*, p.101-109.
16. Lorrai, M., G. M. Sechi, 1995. "Neural nets for modeling rainfall-run transformations," *Wat. Resour. Man.*, 9, p.299-313.
17. Mallat, S. G., 1989. "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(7), p.674-693.
18. Masuda, N. and K. Aihara, "Prediction of chaotic time series with wavelet coefficients," *Electronics and Communications in Japan*, Part 3, 84(6), p1710-1718.

19. McCulloch W. S. and Pitts W., 1943. "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, 5, p.115-133.
20. Minsky M. L. and Papert S. A., 1969. *Perceptrons*. Cambridge. MA:MIT Press.
21. Parios, A. G., O. T. Rais & A. F. Atiya, 2000. "Multi-step-ahead prediction using dynamic recurrent neural networks," *Neural Networks*, vol.13, p.765-786.
22. Rao, R. M., 1998. *Wavelet Transforms, Introduction to Theory and Applications*. Addison:Wesley, USA.
23. Rosenblatt F., 1958. "The perceptron: A probabilistic model for information storage and organization in the brain," *Psych. Rev.*, 65, p.386-408
24. Rumelhart D. E. and McClelland J. L., 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*. Cambridge, MA: MIT Press.
25. Sajikumar, N., and B. S. Thandaveswara, 1992. "A non-linear rainfall-runoff model using an artificial neural network," *Journal of Hydrology*, 216(1), p.32-55.
26. Widrow B. and Hoff M. E. Jr., 1960. "Adaptive switching circuits," *IRE Western Electric Show and Convention Record*, part4, p.96-104
27. Zhang, B. and S. G. Rao, 2000. "Prediction of watershed runoff using Bayesian concepts and modular neural networks," *Water Resources Research*, 36(3), p.753-762.
28. 王志雄，小波複解析法之研究及其應用於感潮河段率定曲線之修正，國立臺灣大學生物環境系統工程學系暨研究所博士論文，1995
29. 王瀚德，小波理論與類神經網路應用於潮汐之預測與補遺，國立中山大學海洋環境及工程學系碩士論文，2001
30. 吳呈懋，利用貝氏法進行地下水數值模式參數檢定，國立臺灣大學土木工程學系碩士論文，1992

31. 李宗穆，小波理論之研究及其應用於水庫流量特性之分析，國立臺灣大學生物環境系統工程學系暨研究所博士論文，1994
32. 李基毓，數值模擬探討井際分溶示蹤試驗估計非水相體飽和度之誤差，國立臺灣大學土木工程學系碩士論文，2004
33. 周建明，水文系統之小波多分辨分析及其應用於流域之洪水預報，國立臺灣大學生物環境系統工程學系暨研究所博士論文，1999
34. 施權展，具混合學習之小波類神經網路完成通訊通道等化器設計，朝陽科技大學網路與通訊研究所碩士論文，2004
35. 洪士林、黃炯憲等，應用小波神經網路於橋樑實測地震反應之損壞偵測，國家地震研究中心研究計畫，2002
36. 胡永國，應用類神經網路推估地下水位洩降所致地層沉陷之研究，屏東科技大學土木工程系碩士論文，1999
37. 胡湘帆，反傳遞模糊類神經網路於流量推估之應用，國立臺灣大學農業工程學系碩士論文，1998
38. 孫建平，類神經網路及其應用於降雨及逕流過程之研究，國立臺灣大學農業工程學系碩士論文，1997
39. 張大元，類神經網路在水庫放流對河川水位增量之研究，中原大學土木工程學系碩士論文，2003
40. 梁晉銘，複合型類神經網路建構集水區水文模式之研究，國立臺灣大學生物環境系統工程學系暨研究所碩士論文，2002
41. 郭建銘，應用多變量統計與類神經網路分析雲林沿海地區地下水水質變化，國立臺灣大學農業工程學研究所碩士論文，1998
42. 陳昶憲、楊朝仲、王益文，類神經網路於烏河流域洪流預報之應用，1996，中華水土保持學報，vol.27，no. 4，p.267-274。

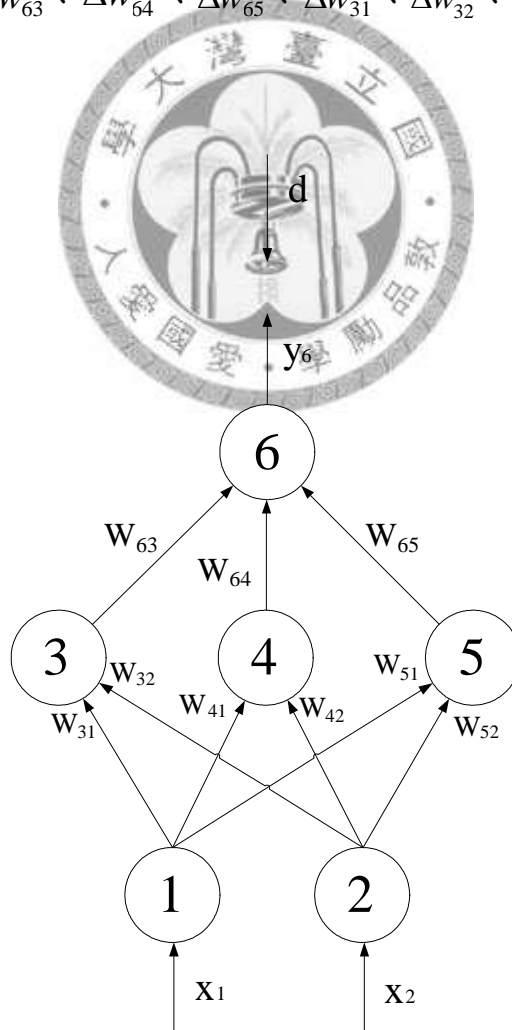
43. 曾國源，以類神經網路架構土石流預警系統之研究，國立臺灣大學生物環境系統工程學系暨研究所博士論文，2003
44. 黃宏志，小波神經網路應用於伺服控制之研究，逢甲大學自動控制工程學系碩士論文，2000
45. 黃智顯，水文時間序列類神經網絡之研究及其應用於流量之預測，國立臺灣大學農業工程學系碩士論文，1997
46. 黃群岳，颱風洪流量之神經網路預測，中原大學土木工程學系碩士論文，2000
47. 楊朝仲，類神經網路在集水區出流量預測之應用，逢甲大學土木及水利工程研究所碩士論文，1997。
48. 廖啟佑，應用類神經網路與小波理論分析地震前地下水位波動，國立臺北科技大學土木與防災研究所碩士論文，2002
49. 蔡奇男，以小波類神經網路實現彩色影像之臉部偵測，朝陽科技大學資訊工程系碩士論文，2004
50. 賴建元，應用類神經網路於屏東平原地下水水位補遺推估，屏東科技大學土木工程系碩士論文，2004
51. 鍾芸菁，應用倒傳遞類神經模式預測山坡地地下水位，國立中興大學土木工程學系碩士論文，2002
52. 王進德，類神經網路與模糊控制理論入門與應用，全華科技，2007
53. 張斐章、張麗秋，類神經網路理論與實務，東華書局，2003
54. 羅華強，類神經網路—MATLAB 的應用，清蔚科技，2001

附錄一 倒傳遞類神經網路公式計算範例

本附錄之範例摘自張斐章、張麗秋等所著之「類神經網路理論與實務」一書（東華書局，2003）之某一範例練習（隨著出版刷次不同，於書中之位置可能有所更動），解題所用之公式可參考該書所述，本論文不再贅述；附上此範例之目的在於舉例解釋倒傳遞類神經之詳細運作過程。

（一）問題描述

下圖為一 BP 類神經網路，若活化函數為 $y = f(net) = 1/(1 + e^{-net})$ ，請求出各權重的修正方程式 Δw_{63} 、 Δw_{64} 、 Δw_{65} 、 Δw_{31} 、 Δw_{32} 、 Δw_{41} 、 Δw_{42} 、 Δw_{51} 、 Δw_{52} 。



(二) 參考解答

權重修正公式為 $\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} = \eta \delta_j^n y_i^{n-1}$,

又依其所在「層位」不同有如下之區分：

$$\left\{ \begin{array}{l} \text{當 } W_{ji} \text{ 處於輸出層與隱藏層之間, } \delta_j^n = (d_j - y_j^n) f'(net_j^n) \\ \text{當 } W_{ji} \text{ 處於隱藏層與隱藏層之間, } \delta_j^n = \left(\sum_k \delta_k^{n+1} w_{kj} \right) f'(net_j^n) \end{array} \right.$$

其中：

$$\begin{cases} y = f(net) = \frac{1}{1 + e^{-net}} \\ y' = f'(net) = \frac{e^{-net}}{(1 + e^{-net})^2} \end{cases}$$

1. 當 $j = 6$, $n = 3$ 時 W_{63} 、 W_{64} 及 W_{65} 處於輸出層及隱藏層間

此時 $\delta_j^n = (d_j - y_j^n) f'(net_j^n)$

$$\begin{aligned} \delta_6^3 &= (d - y_6^3) f'(net_6^3) \\ &= (d - y_6^3) \frac{e^{-net_6^3}}{(1 + e^{-net_6^3})^2} \\ &= [d - f(net_6^3)] \frac{e^{-net_6^3}}{(1 + e^{-net_6^3})^2} \\ &= \left[d - \frac{1}{1 + e^{-net_6^3}} \right] \frac{e^{-net_6^3}}{(1 + e^{-net_6^3})^2} \end{aligned}$$

又 $net_6^3 = W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2$

$$\text{即 } \delta_6^3 = \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2}$$

$$\text{其中 } \begin{cases} y_3^2 = f(net_3^2) = \frac{1}{1 + e^{-net_3^2}} = \frac{1}{1 + e^{-(x_1W_{31} + x_2W_{32})}} , & net_3^2 = x_1W_{31} + x_2W_{32} \\ y_4^2 = f(net_4^2) = \frac{1}{1 + e^{-net_4^2}} = \frac{1}{1 + e^{-(x_1W_{41} + x_2W_{42})}} , & net_4^2 = x_1W_{41} + x_2W_{42} \\ y_5^2 = f(net_5^2) = \frac{1}{1 + e^{-net_5^2}} = \frac{1}{1 + e^{-(x_1W_{51} + x_2W_{52})}} , & net_5^2 = x_1W_{51} + x_2W_{52} \end{cases}$$

$$\Delta w_{63} = \eta \delta_6^3 y_3^2 = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} y_3^2$$

$$\Delta w_{64} = \eta \delta_6^3 y_4^2 = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} y_4^2$$

$$\Delta w_{65} = \eta \delta_6^3 y_5^2 = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} y_5^2$$

2. 當 $j = 3$, $n = 2$ 時 W_{32} 及 W_{31} 處於隱藏層及隱藏層間

$$\text{此時 } \delta_j^n = \left(\sum_k \delta_k^{n+1} w_{kj} \right) f'(net_j^n)$$

$$\delta_3^2 = (\delta_6^3 W_{63}) f'(net_3^2)$$

$$= (\delta_6^3 W_{63}) \frac{e^{-net_3^2}}{(1 + e^{-net_3^2})^2}$$

$$= \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{63} \frac{e^{-(x_1 W_{31} + x_2 W_{32})}}{[1 + e^{-(x_1 W_{31} + x_2 W_{32})}]^2}$$

$$\Delta W_{31} = \eta \delta_3^2 x_1^1$$

$$\Delta W_{31} = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{63} \frac{e^{-(x_1 W_{31} + x_2 W_{32})}}{[1 + e^{-(x_1 W_{31} + x_2 W_{32})}]^2} x_1^1$$

$$\Delta W_{32} = \eta \delta_3^2 x_2^1$$

$$\Delta W_{32} = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{63} \frac{e^{-(x_1 W_{31} + x_2 W_{32})}}{[1 + e^{-(x_1 W_{31} + x_2 W_{32})}]^2} x_2^1$$

$$\text{其中 } \begin{cases} y_3^2 = f(net_3^2) = \frac{1}{1 + e^{-net_3^2}} = \frac{1}{1 + e^{-(x_1 W_{31} + x_2 W_{32})}} , & net_3^2 = x_1 W_{31} + x_2 W_{32} \\ y_4^2 = f(net_4^2) = \frac{1}{1 + e^{-net_4^2}} = \frac{1}{1 + e^{-(x_1 W_{41} + x_2 W_{42})}} , & net_4^2 = x_1 W_{41} + x_2 W_{42} \\ y_5^2 = f(net_5^2) = \frac{1}{1 + e^{-net_5^2}} = \frac{1}{1 + e^{-(x_1 W_{51} + x_2 W_{52})}} , & net_5^2 = x_1 W_{51} + x_2 W_{52} \end{cases}$$

3. 當 $j = 4$, $n = 2$ 時 W_{42} 及 W_{41} 處於隱藏層及隱藏層間

$$\text{此時 } \delta_j^n = \left(\sum_k \delta_k^{n+1} w_{kj} \right) f'(net_j^n)$$

$$\begin{aligned} \delta_4^2 &= (\delta_6^3 W_{64}) f'(net_4^2) \\ &= (\delta_6^3 W_{64}) \frac{e^{-net_4^2}}{(1 + e^{-net_4^2})^2} \\ &= \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{64} \frac{e^{-(x_1 W_{41} + x_2 W_{42})}}{[1 + e^{-(x_1 W_{41} + x_2 W_{42})}]^2} \end{aligned}$$

$$\Delta W_{41} = \eta \delta_4^2 x_1^1$$

$$\Delta W_{41} = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{64} \frac{e^{-(x_1 W_{41} + x_2 W_{42})}}{[1 + e^{-(x_1 W_{41} + x_2 W_{42})}]^2} x_1^1$$

$$\Delta W_{42} = \eta \delta_4^2 x_2^1$$

$$\Delta W_{42} = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{64} \frac{e^{-(x_1 W_{41} + x_2 W_{42})}}{[1 + e^{-(x_1 W_{41} + x_2 W_{42})}]^2} x_2^1$$

$$\text{其中 } \begin{cases} y_3^2 = f(net_3^2) = \frac{1}{1 + e^{-net_3^2}} = \frac{1}{1 + e^{-(x_1 W_{31} + x_2 W_{32})}} , & net_3^2 = x_1 W_{31} + x_2 W_{32} \\ y_4^2 = f(net_4^2) = \frac{1}{1 + e^{-net_4^2}} = \frac{1}{1 + e^{-(x_1 W_{41} + x_2 W_{42})}} , & net_4^2 = x_1 W_{41} + x_2 W_{42} \\ y_5^2 = f(net_5^2) = \frac{1}{1 + e^{-net_5^2}} = \frac{1}{1 + e^{-(x_1 W_{51} + x_2 W_{52})}} , & net_5^2 = x_1 W_{51} + x_2 W_{52} \end{cases}$$

4. 當 $j = 5$, $n = 2$ 時 W_{52} 及 W_{51} 處於隱藏層及隱藏層間

$$\text{此時 } \delta_j^n = \left(\sum_k \delta_k^{n+1} w_{kj} \right) f'(net_j^n)$$

$$\begin{aligned} \delta_5^2 &= (\delta_6^3 W_{65}) f'(net_5^2) \\ &= (\delta_6^3 W_{65}) \frac{e^{-net_5^2}}{(1 + e^{-net_5^2})^2} \\ &= \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{65} \frac{e^{-(x_1 W_{51} + x_2 W_{52})}}{[1 + e^{-(x_1 W_{51} + x_2 W_{52})}]^2} \end{aligned}$$

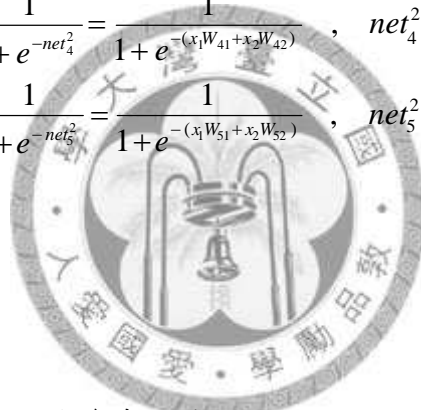
$$\Delta W_{51} = \eta \delta_5^2 x_1^1$$

$$\Delta W_{51} = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{65} \frac{e^{-(x_1W_{51} + x_2W_{52})}}{[1 + e^{-(x_1W_{51} + x_2W_{52})}]^2} x_1^1$$

$$\Delta W_{52} = \eta \delta_5^2 x_2^1$$

$$\Delta W_{52} = \eta \left[d - \frac{1}{1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}} \right] \frac{e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}}{[1 + e^{-(W_{63}y_3^2 + W_{64}y_4^2 + W_{65}y_5^2)}]^2} W_{65} \frac{e^{-(x_1W_{51} + x_2W_{52})}}{[1 + e^{-(x_1W_{51} + x_2W_{52})}]^2} x_2^1$$

$$\text{其中} \begin{cases} y_3^2 = f(net_3^2) = \frac{1}{1 + e^{-net_3^2}} = \frac{1}{1 + e^{-(x_1W_{31} + x_2W_{32})}} , & net_3^2 = x_1W_{31} + x_2W_{32} \\ y_4^2 = f(net_4^2) = \frac{1}{1 + e^{-net_4^2}} = \frac{1}{1 + e^{-(x_1W_{41} + x_2W_{42})}} , & net_4^2 = x_1W_{41} + x_2W_{42} \\ y_5^2 = f(net_5^2) = \frac{1}{1 + e^{-net_5^2}} = \frac{1}{1 + e^{-(x_1W_{51} + x_2W_{52})}} , & net_5^2 = x_1W_{51} + x_2W_{52} \end{cases}$$



以上即為本附錄之範例題目與參考解答，唯一要特別提出的是，此範例中隱藏層與輸出層之活化函數皆為 $y = f(net) = 1/(1 + e^{-net})$ ，但本研究中只有隱藏層採用雙曲正切函數，輸出層活化函數改為線性函數；兩者雖然有些許不同，但只有使得最後答案之多項式形式不一樣，並不會影響其對於倒傳遞類神經網路運作方式之詮釋。

附錄二 小波濾除權重範例

假設一待分析序列為 $[10\ 10\ 10\ -10\ 10\ 15\ 20\ 15\ 10\ 10\ 15\ 19\ -8\ 10\ 10]$ ，其原始波型如下圖1所示，放入小波模式裡過濾之後，被分為A1（下圖2）與D1（下圖3）兩部份，其中A1代表平均值，D1代表擾動項；根據小波的原理，需要進一步分析者是D1，看看在這個D1裡面有什麼地方是振盪幅度大的，就是重要的訊號，振盪幅度大的點其數值予以保留，不夠大的點其數值就予以捨棄然後補零，而取捨的標準就是本論文中所謂的門檻值（本論文取0.6）；選取方法請見3.5.1節所述，最後選取出來的結果如下圖4所示。

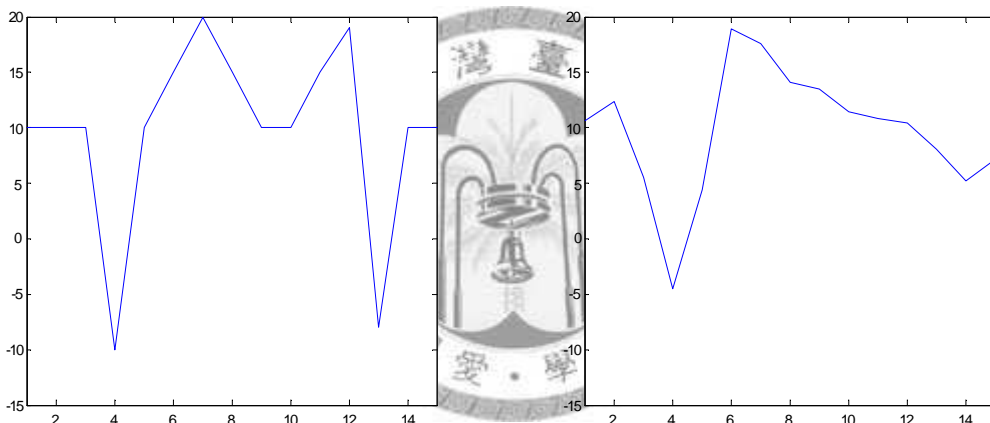


圖 1 原始訊號

圖 2 平均值 A1

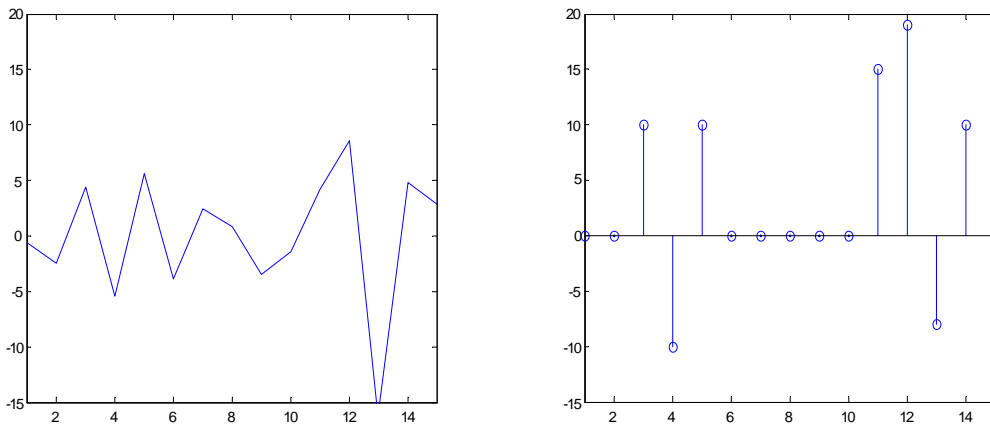


圖 3 擾動項 D1

圖 4 小波分析後留下之點