第四章 研究設計與結果分析

4-1. 研究設計

本研究係探討如何從氣候與農產品兩異質性之資料庫,發掘彼此資料之關聯程度,並建構一預測模型,以利日後相關預測之依據。故以「監督式知識發掘」(supervised knowledge discovery)為基本研究主架構。「監督式知識發掘」是從既有的資料中去發掘出有意義資訊,並以此結果,解釋過去已發生的事件,以作為未來的預測。由於本研究所分析的農產品資料,因其有生長期之限制,故具有時間序列特性。所以本研究選擇倒傳遞類神經網路(BPN)作為資料探勘的工具。並藉由建置一個氣候-產地蔬菜產量模型,以預測產地的氣候變因對蔬菜產量收成的影響。

在建構輸入層-隱藏層-輸出層三層式的 BPN 類神經網路架構前, 必須先完成下述前置作業:

- 將資料分成訓練範例與測試範例
- 先以無隱藏層的雙層 BPN 類神經網路架構,測試資料是否有線性相關。
- 設定隱藏層層數與每層處理單元數學習法則
- 漸次輸入相關資料,以進行網路學習。並觀察學習結果,再決定是否增減隱藏層層數或每層處理單元數

故在整個類神經網路的建置的過程中,本研究中會先以「一層隱藏層」、「二層隱藏層」的方式做對照,以分析出何種型式的類神經網路,是建置本研究的最佳預測模型。

4-2. 資料的選取

一、農產品資料的選取

本研究的主要研究重點,係在探討如何利用「資料探勘」技術, 於時間序列的歷史資料中,發掘出有意義的知識或資訊,以作為日後 預測的依據。因此,在時間序列的資料種類中,由於農產品批發市場 每日的交易價格與交易數量具有時間上的連續性,故作為本研究基本 的資料庫。

有關農產品的相關資料的來源,例如:蔬菜或水果的種類、產地來源、批發市價、零售市價、成交最高價、最低價、平均價與成交數量等,皆從以下兩個網站所公佈的資料中擷取:

- 1. 行政院農業委員會農產品行情交易網站 http://163.29.73.197/
- 2. 行政院農業委員會農業產銷班資訊服務網 http://farm.coa.gov.tw/

由於本研究主要是以時間序列的資料為主,因此,在農產品標的的挑選上,必須符合:

- 1. 該項農產品必須一年四季皆有生產,方符合時間序列所要求的連續性資料。
- 2. 該項農產品須為民生消費大宗,其交易量具有決定價格之特性。
- 3. 該項農產品之產地產量,容易受到其他變數(例如:天氣)的影響。

根據台北第一、第二果菜交易市場的交易紀錄發現,由於水果的 栽種方式與採收狀況,係根據季節性的不同,而生產符合當時的節令 水果。因此,在一年中某特定的片段時間內,才是特定水果的產季;這種不具時間連續性的資料,並不符合時間序列的要求。

在蔬菜方面,以交易量而言,根據[表 4-1]所顯示,民國 88 年度台灣主要消費地批發市場蔬菜交易價格與數量統計顯示,台北第一與第二果菜市場蔬菜總交易量就達 374681.5 公噸,佔全年度台灣地區總市場交易量 873100.3 公噸的 42.91%,佔四成多。又根據[表 4-2]及[表 4-3]的結果顯示,民國 88 年度台北第一、第二果菜交易市場甘藍菜的年度交易量占葉菜類總成交量達三成,為果菜批發市場中交易量最大宗之蔬菜。葉菜類蔬菜,容易受到氣象變因之影響,例如:降雨量、溫度、日照率和風速等影響,故本研究根據台北第一與第二果菜市場蔬果交易狀況,選定甘藍(初秋品種)作為本研究之標的蔬菜。

又根據第二章文獻探討得知,甘藍(初秋)主要產地集中於彰化與雲林兩縣。根據[表 4-4]得知位於彰化縣的溪湖果菜批發市場與雲林縣的西螺果菜批發市場,在民國 88 年度台灣地區甘藍菜的總交易量上,分別佔有 55. 78%與 22. 76%,位居交易量第一、第二名。但在資料的分析與篩選過程中,發現溪湖果菜批發市場甘藍菜的歷年交易量紀錄中,有缺項的情況,較不適用於時間序列的資料分析。故本研究決定以西螺果菜批發市場的甘藍菜交易量作為具有時間序列性質的歷史資料庫,交易時間由 87/01/01 至 89/04/30。

表 4-1、民國 88 年度台灣主要消費地批發市場蔬菜交易價格與數量

市場名稱	平均價(元)	交易量(公噸)	產量百分比
台北二	19.4	60936.8	6.98%
台北一	21.4	313744.7	35.93%
台中市	16.0	17657.2	2.02%
永靖鄉	15.5	27526.6	3.15%
溪湖鎭	20.3	56366.3	6.46%
南投市	17.7	3934.6	0.45%
西螺鎭	17.6	225889.9	25.87%
高雄市	20.0	60657.4	6.95%
鳳山市	11.2	14022.9	1.61%
屏東市	12.6	72464.8	8.30%
台東市	25.6	8911.8	1.02%
花蓮市	16.2	10987.3	1.26%
產量		873100.3	100.00%

資料來源:本研究整理

表 4-2、台北第一果市交易市場 88 年度交易量前三名蔬菜月交易量一覽表 (單位:公斤)

								(-	PIM·A/I/
月份	甘藍(總)	交易比例	包心白菜 (總)	交易比例	青江白菜	交易比例	葉菜類 總交易量	葉菜類比例	總成交量
88/01	3,795,696	30.83%	3,442,882	27.96%	450,901	3.66%	12,312,660	38.12%	32,297,323
88/02	723,750	27.99%	550,763	21.30%	122,673	4.74%	2,585,842	9.40%	27,494,887
88/03	3,079,768	30.73%	2,170,092	21.66%	496,764	4.96%	10,020,825	34.98%	28,649,952
88/04	2,192,852	31.69%	1,101,778	15.92%	449,154	6.49%	6,918,658	24.22%	28,565,575
88/05	3,141,736	36.79%	1,312,014	15.36%	501,366	5.87%	8,539,906	30.91%	27,626,323
88/06	2,811,397	33.09%	1,388,933	16.35%	563,702	6.63%	8,496,569	31.56%	26,922,618
88/07	3,093,629	36.55%	1,817,287	21.47%	425,747	5.03%	8,463,885	30.90%	27,391,196
88/08	2,186,829	44.63%	909,056	18.55%	156,550	3.20%	4,899,387	20.34%	24,081,541
88/09	1,926,837	29.53%	1,344,162	20.60%	490,332	7.52%	6,524,288	25.63%	25,452,287
88/10	3,358,394	31.80%	2,753,660	26.07%	507,173	4.80%	10,561,417	37.53%	28,139,449
88/11	3,329,282	31.67%	2,357,425	22.42%	564,236	5.37%	10,513,657	35.94%	29,250,975
88/12	3,830,905	32.13%	3,116,221	26.13%	531,403	4.46%	11,924,451	39.31%	30,334,715
	!交易 比例	33.12%				總計	101,761,545	30.27%	336,206,841

資料來源:本研究整理

表 4-3.台北第二果市交易市場 88 年度交易量前三名蔬菜月交易量一覽表 (單位:公斤)

月份	甘藍(總)	交易比例	包心白菜 (總)	交易比例	青江白菜	交易比例	葉菜類 總交易量	葉菜類比例	總成交量
88/01	949,640	31.93%	690,298	23.21%	95,545	3.21%	2,974,574	41.88%	7,101,960
88/02	203,137	30.76%	122,515	18.55%	30,497	4.62%	660,357	11.06%	5,971,542
88/03	768,175	32.19%	444,259	18.62%	116,227	4.87%	2,386,022	38.06%	6,269,398
88/04	639,526	34.99%	215,525	11.79%	133,982	7.33%	1,827,887	28.82%	6,342,539
88/05	720,820	35.38%	264,407	12.98%	120,170	5.90%	2,037,152	33.63%	6,057,159
88/06	639,746	33.35%	236,955	12.35%	128,411	6.69%	1,918,183	34.07%	5,629,602
88/07	659,982	37.22%	312,309	17.61%	96,454	5.44%	1,773,187	29.57%	5,995,711
88/08	447,078	47.39%	120,291	12.75%	24897	2.64%	943,450	19.01%	4,962,413
88/09	363,470	27.42%	208,061	15.69%	111695	8.42%	1,325,787	24.83%	5,339,133
88/10	718,550	34.46%	426,306	20.45%	94262	4.52%	2,085,121	33.60%	6,205,101
88/11	824,465	32.46%	453,314	17.85%	122751	4.83%	2,539,621	39.13%	6,490,528
88/12	817,250	33.53%	502,961	20.64%	108115	4.44%	2,437,377	36.49%	6,679,447
	交易 比例	34.26%				總計	22,908,718	31.36%	73,044,533

資料來源:本研究整理

表 4-4、民國 88 年度台灣主要消費地批發市場甘藍菜交易量分析

(單位:公噸)

年/月	台北	台中	溪湖	西螺	高雄
88/01	5438.0	318.7	2695.0	7328.0	1017.0
88/02	4151.0	267.6	1943.0	5559	968.0
88/03	4372.0	201.0	1996.0	4742	981.0
88/04	4270.0	217.4	1313.0	4203	972.0
88/05	4522.0	229.6	990.0	4517	1246.0
88/06	4144.0	270.0	705.0	5578	988.0
88/07	4306.0	388.0	199.0	6161	1020.0
88/08	3967.0	394.0	91.0	5927	739.0
88/09	3064.0	266.0	159585.0	5703	533.0
88/10	4055.0	320.0	960.0	6224	653.0
88/11	5091.0	354.0	2394.0	7481	440.0
88/12	5309.0	372.0	810.0	7451	985.0
總計	52689.0	3598.3	173681.0	70874.0	10542.0
比例	16.92%	1.16%	55.78%	22.76%	3.39%
88 年	度台灣地區甘藍	蓝菜總交易量			311384.3

資料來源:本研究整理

二、氣象資料的選取

根據文獻探討得知,產地蔬菜在生長的過程中,會受到以下氣象因素的影響:溫度、溼度、蒸發皿中之蒸發量、日照和降雨量[洪英琳,1998]。本研究採用中央氣象局所提供的逐日地面氣象資料表(A、B表,請參看[圖 4-1]與[圖 4-2])中的:

- 1. 平均氣溫
- 2. 温差
- 3. 相對溼度
- 4. 平均風速
- 5. 降水量
- 6. 日照率

作為影響產地蔬菜生長的氣象變異因素,資料觀測時間是 1998/01/01-2000/03/31。

1661	報数		記	微微		86		88	1	1											291				
					16.4	50	-				_														
June 12,		十冊	*	超小	大車	82	-	22	030	708															
	0.1 mm	一一	K	李	大事品	78	_	8	0045	3847	代題	6	2	Ξ	12	13	14	15	16						
		盐		额	0.1	75	-	77	024	081															
	降木量		山		g	1	_	74	0045	900E	第一 関心名称	SSW	SW	WSW	W	WNW	ΜN	NNN	z						
	- 世	m		提	太。	89	_	20	022	033	が国る関連の名類														
	蒸發量	0.1mm		摄	Æ	65		19	084 012 022	020	題	南南西	便	西南西	西	西北西	西北	北北西	÷						
		展	陞	14	T	62	-	R		090	記録			30											
	9	K		恒	:	8	-	19	12	10		-	2	3	4	45	9	-	00				8		
	91~1	盛		敪		57	Ξ	59	057	060	N. Carrier and St.	m										-			
	lm/s	K		10	*	-		99	5 12	4 10	向名稱	NNE	S	ENE	Э	ESE	SS	SSE	S				E SECTION		
	風, 0.	極		殷		52	_	54	088 031 085	5 10	風向	上北東	11	東北東		東南東	200	南南東		2		100			
		各		型		49	_	2	3 03	0.5		#	東北	张	ङ	展	東南	車	佢			Adverse	星	_	_
	鱼	畢		din		46	-	48	380	1 075	_	_	_		_	_		_					E E	7.8	101
	見	部分	継	對縣	2. %	43	_	45	024	047	Sign)											0000	6,666	88	
Š.,		Ŷ	*	粃		9	_	42	146 052	034	(Negative Sign)	TO:	-	×	_	×	z	0	d.	0	25	100 5 12		86-87	00 00
(NEW)	0.17		謡	南		37	-	39		05.3	艺外	r) or										11-15	CH III	500	30
5	期期		順	框		34	_	36	198	01P	負値字	100		. 9								100 (1) 10	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	26	>
	*		H	型		3	-	33	169	02.3	Sign)			- 40								111.00	E E	8	0,0
			嶝	岜		26	-	30	10156	467550 1997 01 03 J 06393 06566 06289 02 J 01 P 05 J 034 044 075 055 104 10 090 10 060 020 03 J 006 P 08 J 3847 708	字值 (Digit Values) 正值字 (Positive Sign) 負値字(Negative S	? or 0	Aorl	B or 2	Cor 3	Dor 4	E or 5	For 6	G or 7	H or 8	6 JO		で 序/人用 現代大大大国 超過 995.5mm 質	7.82	0.04
	0.1 hPa		暗	施		21	_	25	10218	9999	正值字 (Positive	or or			Ĭ				Ĭ			· 计数据 电影性 多彩 日本版本。	- XX	9-1	9
KINGHAN MA	紙票			-						93 00	(S) IF						-			.0.	-	100	200	17	22
				红		91	_	20	J 10181	063	字值 (Digit Values)											A 2011	T ME	P30121	CC I UDI
	1/4	TE.	盟	쁆	_	13 15	-			3 J	(Digi	0	-	2	m	4	5	9	1	00	6	- 1			
	E					=	_	27	090	010	字值								9			1 1	Š.		17.5
	擅		計			7	_	0	166	166												68.00	K	報位達	X:脚除。19: 個人完
	-					-		-	21 1	50 1												200	N 55K 31	おは	
	丽	智	S K	回		-		9	466921 1997 06 01	4675												6		3. 新獎格式欄位對照	X-III
	臣	ш	矽	筵			翻位		報酬																

圖 4-1、逐日地面氣象資料-A 表 資料來源:中央氣象局

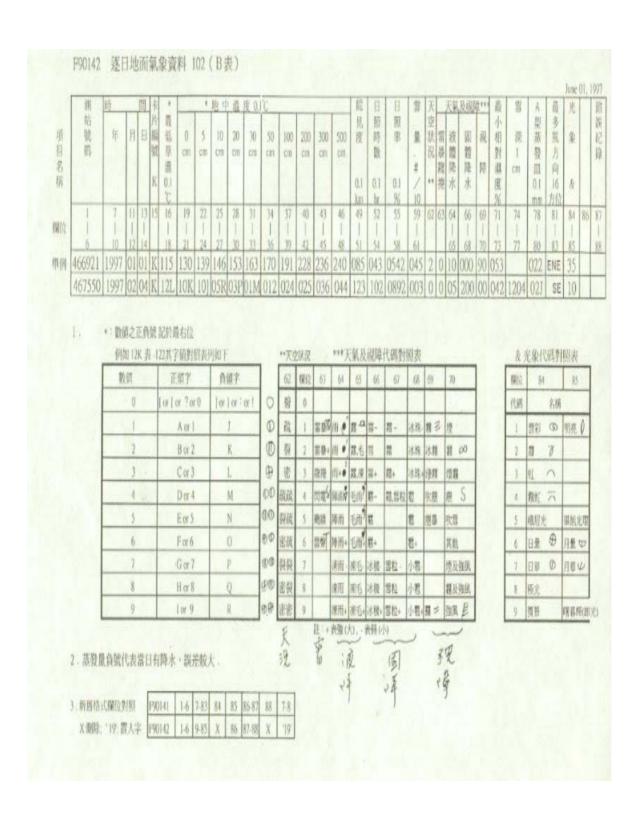


圖 4-2、逐日地面氣象資料-B 表 資料來源:中央氣象局

4-3. 研究建置步驟

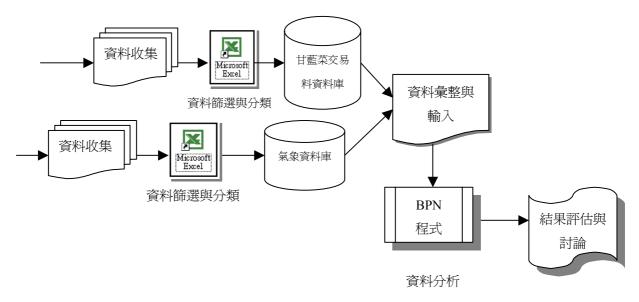


圖 4-3、本研究實體架構圖 資料來源:本研究整理

本研究之建置步驟如下:

一、資料收集:

本研究所研究之基本資料來源為:

- 氣候資料:根據中央氣象局所提供的逐日地面氣象資料 (A、B表)為基準。
- 農產品資料:以行政院農委會農產品交易網站所公佈之蔬菜 交易量為基準。

二、資料彙整與處理:

本階段係在將所收集之資料,依照研究設計所需,分別加以彙整、刪減與補值處理。

■ 氣候資料:將逐日地面氣象資料 A、B 兩表彙整成一表後, 僅擷取本研究所需的欄位資料,如:平均氣溫、溫差、相對 溼度、平均風速、降水量和日照率。本研究以台中、彰化、 雲林及嘉義等四縣市之地面測候站之測候資料為主。

■ 農產品資料:由前述農產品資料整理得知,經刪減與整合後, 本研究則以西螺果菜批發市場的甘藍菜交易量(依照雲林縣 西螺果菜批發市場之交易紀錄),作為另一研究用之主要資 料。

三、類神經網路建置:

有鑑於 Matlab 提供許多的函式庫及工具箱,以利使用者之程式之開發。故本研究以 Matlab 為主要 BPN 程式之開發工具。並利用 Matlab 所撰寫的 BPN 程式來訓練及測試模型。

在 BPN 相關設定上:

- 輸入層的輸入因子:即平均氣溫、溫差、相對溼度、平均風速、 降水量和日照率等氣候因素。
- 輸出層的目標值:以西螺果菜批發市場的甘藍菜交易量為目標 值。
- 隱藏層數的選擇:一層或兩層
- 隱藏層處理單元:採用平均法,即(輸入處理單元數+輸出處理單元數)/2。
- 網路架構:由於甘藍菜的生長期約為100天,而產地的氣候因素係對生長期間的甘藍菜才有影響效力。換言之,在產地批發市場中當日的交易量,應與成交日當日前推三個月內每一日氣候因素作相關影響分析。故本研究以輸出變數-甘藍菜成交量當日,前推三個月,並以週為單位(即13週),對每週中所有輸入變數(即氣候因素的平均氣溫、溫差、相對溼度、平均風速、降水量和日照率),取其極大值、極小值和差異值(即極大值與

極小值之差)作為總輸入變數。以此研究設計,將會使實驗的環境變數更接近真實的狀況。但亦會造成輸入變數非常龐大,因為輸出變數-成交量當日要對映之前13週所有的輸入變數的極大值、極小值和差異值,加上原有的六個變數,總共將有6(個變數)*3(個值)*13(週)+6(原有變數)=240個輸入變數。

本研究之網路架構採:

- 隱藏層一層時:隱藏層之處理單元採輸入值之一半:240/2=120,即 240-120-1 架構。
- 隱藏層二層時:第二隱藏層之處理單元再採第一隱藏 層處理單元輸入值之一半: 即 240-120-60-1 架構。
- 樣本:按照西螺果菜批發市場的甘藍菜 87/04/01-87/11/30 等日交易量為訓練樣本,共計 610 筆。而以 87/12/01-88/04/30 等日交易量為測試樣本,共計 151 筆。
- 亂數種子:固定為一個常數值,此數值的差異對訓練的結果的 影響很小。
- 學習速率:使用 default 值(1.04),研究中曾以 0.99 及 1.05 試做比較,但效果不明顯故使用 default 值。
- 學習循環數目:隱藏層一層,訓練次數 100-1000 次。 隱藏層兩層,訓練次數 100-8500 次。
- 其它未定義的參數值:皆按照 Matlab 類神經工具箱中的 default 值,不另加說明。

綜合以上的說明,整理如[表 4-5]、[表 4-6]及參考[圖 4-4]。

表 4-5、多變量 BPN 類神經網路架構設計表

處理 單元	輸出變數	輸入變數				輸入變數					輸入變數			輸入變數				輸入變數				輸入變數			
原變數	交易量	平均氣溫			四山	溫差					對	溼厚	妄	平均風速					降力	火量		日照率			
增列 變數	無	原變數	極大値	極小値	差距値	原變數	極大値	極小値	差距値	原變數	極大値	極小値	差距値	原變數	極大値	極小値	差距值	原變數	極大値	極小値	差距值	原變數	極大値	極小値	差距値
訓練	87/04/01	87/01/01			1	87/01/01				87/01/01			87/01/01			87/01/01			87/01/01						
範例	88/11/30	8	8/0	8/3	1	88/08/31				88/08/31			88/08/31			88/08/31			88/08/31						
測試	88/12/01	8	8/0	9/0	1	8	8/0	9/0	8	88/09/01			88/09/01			88/09/01			88/09/01						
範例	89/04/30	8	9/0	3/3	1	8	9/0	3/3	1	8	9/0	3/3	1	8	89/03/31			89/03/31				89/03/31			
訓練	範例筆數						-				-		61	10		-			-						
測試	範例筆數		151																						

網路架構:240-120-1

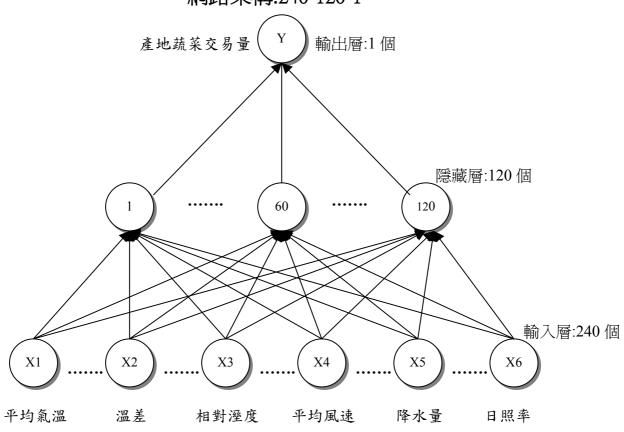


圖 4-4、本研究 BPN 網路架構

表 4-6、多變量 BPN 類神經網路訓練一覽表

網路模式	網路架構	學習循環	測試週期	精準度	亂數種子
BPN	240-120-1	100-1000	50	1E-6	192736547
DIN	240-120-60-1	100-8500	100	1E-6	192736547

4-4. 結果分析

本研究以西螺果菜批發市場甘藍菜交易量為目標值(即圖 4-5 至圖 4-16 中的原始資料曲線),與經執行 151 筆測試資料後所產生的曲線相比對,發現以下結果:

一、隱藏層設為一層:

- 訓練次數在100次時,訓練至67-73筆測試資料後,測試結果才漸趨符合原始值。參考[圖4-5]。
- 2. 訓練次數到達 1000 次以上時,整段訓練值曲線與原始值曲線接近。參考[圖 4-9]。

二、隱藏層設為二層:

- 1. 訓練次數在700次以後,整段訓練值曲線才與原始值曲線接近。參考[圖4-12]。
- 2. 訓練次數達到 3000 次時,整段訓練值曲線已與原始值曲線相當接近。參考[圖 4-14]。
- 3. 但訓練次數達到 8500 次時,整段訓練值曲線卻開始呈現上下震盪之走向,研判是訓練過度(Overlearning)所造成。參考[圖 4-16]。

根據上述結果觀察發現,本研究所設計之 BPN 模型,在適當的訓練次數下,具有不錯的預測效果。

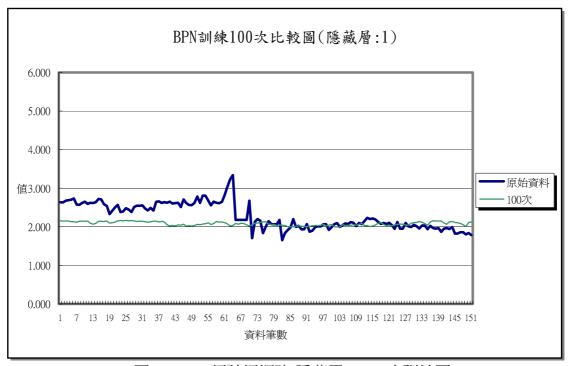


圖 4-5.BPN 類神經網路(隱藏層:1)100 次訓練圖

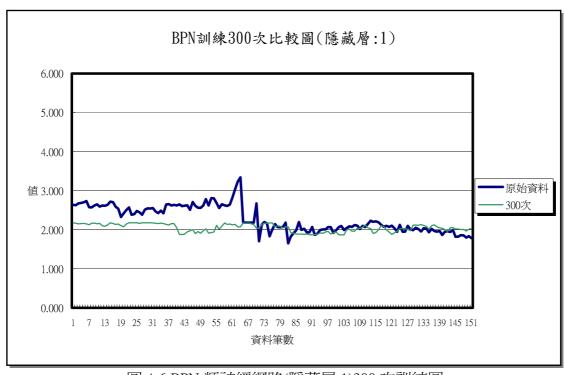


圖 4-6.BPN 類神經網路(隱藏層:1)300 次訓練圖

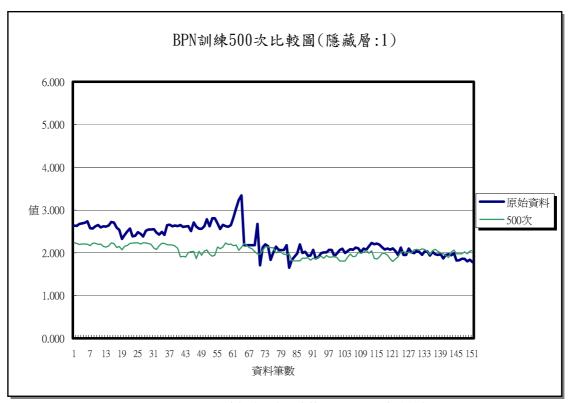


圖 4-7.BPN 類神經網路(隱藏層:1)500 次訓練圖

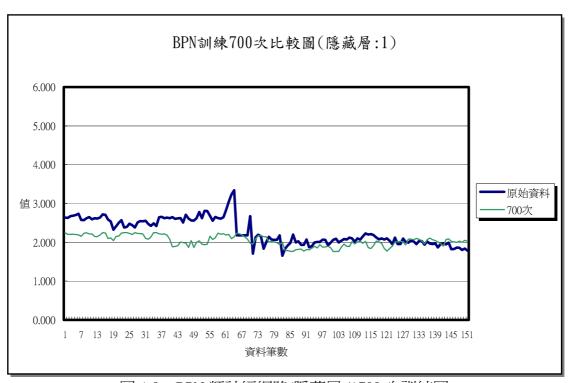


圖 4-8、BPN 類神經網路(隱藏層:1)700 次訓練圖

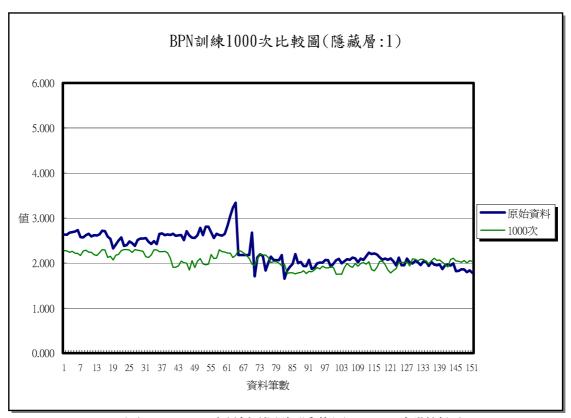


圖 4-9、BPN 類神經網路(隱藏層:1)1000 次訓練圖

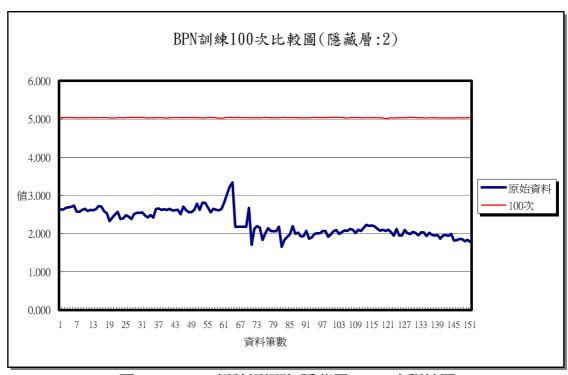


圖 4-10、BPN 類神經網路(隱藏層:2)100 次訓練圖

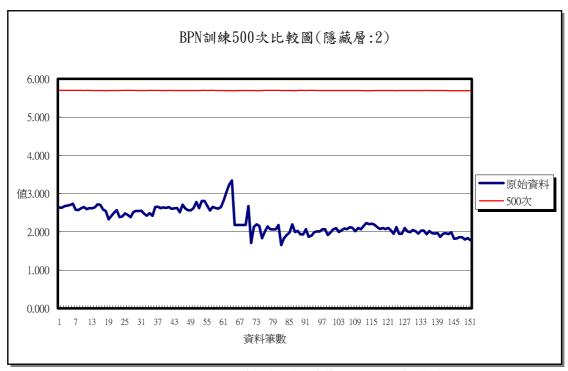


圖 4-11、BPN 類神經網路(隱藏層:2)500 次訓練圖

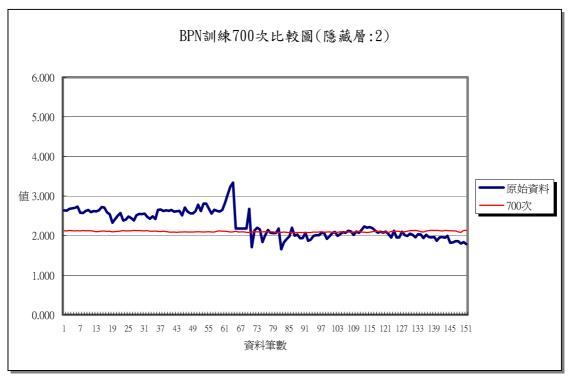


圖 4-12、BPN 類神經網路(隱藏層:2)700 次訓練圖

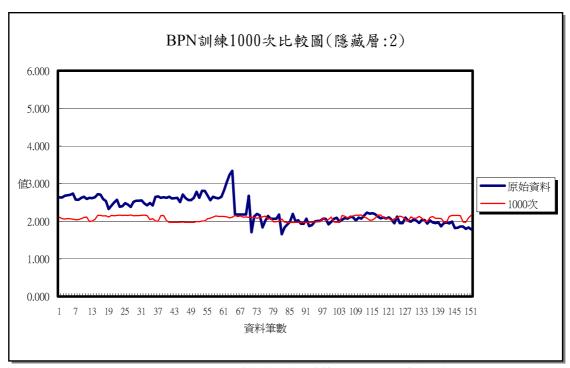


圖 4-13、BPN 類神經網路(隱藏層:2)1000 次訓練圖

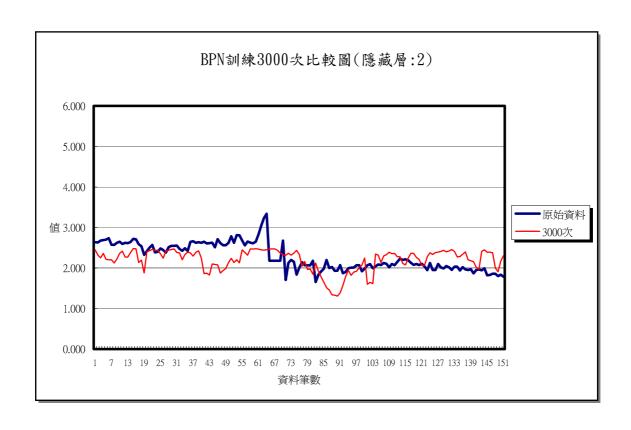


圖 4-14、BPN 類神經網路(隱藏層:2)3000 次訓練圖

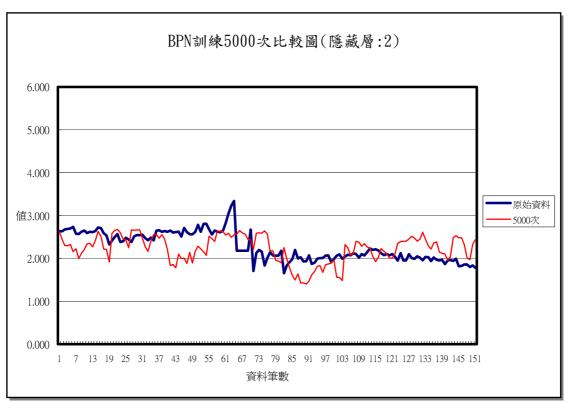


圖 4-15、BPN 類神經網路(隱藏層:2)5000 次訓練圖

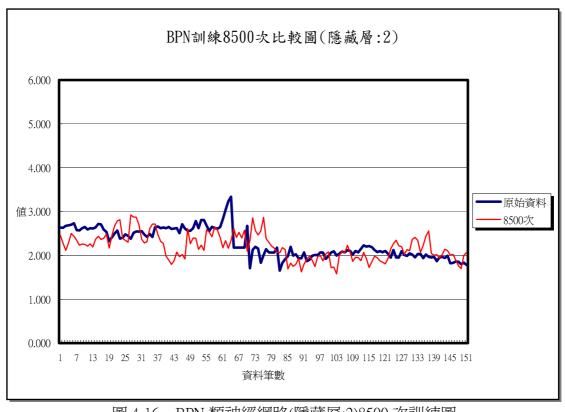


圖 4-16、BPN 類神經網路(隱藏層:2)8500 次訓練圖