

倒傳遞網路與基因演算法於健保投保金額 申報稽核之研究

黃有評
國立臺北科技大學
電機工程系
yphuang@ntut.edu.tw

鄭文進
大同大學
資訊工程研究所
gary.jin@mas.hinet.net

洪茂盛
國立臺北教育大學
數學暨資訊教育系
horng@tea.ntue.edu.tw

摘要

全民健康保險係一種自助互助危險分擔之社會保險制度，為達此目的，必須全體國民均依規定參加保險，繳交保險費，以享受適當的醫療照顧。因此，如何落實全民強制納保辦理投保手續，及防範杜絕違法之情事發生，以期建立永續經營的全民健康保險制度是健保局所努力的目標，健保局一貫秉持以公平正義的態度處理各種不同型態的問題，收入與支出之間的平衡點往往也影響到人民就醫的權利與品質，如何有效掌控此平衡點往往是決策者努力的方向。本研究主要是運用倒傳遞網路與基因演算法技術，從投保金額申報異常資料中，建構出可用以預測投保金額申報異常之分類模型，有效協助稽核人員掌握投保金額申報異常之投保單位，減少因稽核人員誤判造成投保單位必須執行申覆之煩瑣作業程序。實驗結果顯示，利用所提方法可提昇正確辨識率達到94.24%，比一般稽核人員判別的正确率84.63%高出許多。

關鍵詞：倒傳遞網路、基因演算法、投保金額。

1. 序論

1.1 研究背景與動機

健保局從民國84年開辦以來，全民健保財務是採行自給自足及隨收隨付原則，財源主要來自保險費。但近年來收入的平均成長低於醫療費用的平均成長，95年保險收入為3,819億元，保險支出（醫療費用）為3,822億元，然而醫療費用年平均成長率高達5.54%，而保險收入的年平均成長率只有4.70%，出現入不敷出的情形，未來唯有健全健保財務，才能讓健保永續經營，為有效改善健保支出超過收入的情況，近年來竭盡所能陸續實施各項開源與節流措施，為困窘的健保財務補血，因此健保得以持續至今，開源與節流的財務挹注功不可沒，解決龐大的健保財務危機，開源與節流必須並重，並符合社會的公平性；94年間再繼續推出「健保改革多元微調方案」措施之一，包括調

高投保金額上限，由87,600元調整為131,700元，因此許多投保單位投保金額以多報少的情況日趨嚴重。

健保財務一貫秉持以公平正義的態度處理各種不同型態的問題，在資源有限的前提下如何發揮最大效益是其所追求的目標，收入與支出之間的平衡點往往也影響到人民就醫的權利與品質，如何有效掌控此平衡點往往是決策者努力的方向，其收入面的收繳率亦牽動支出面醫療院所服務的品質，其投保單位投保金額的申報稽核更甚為重要，如何將投保金額申報稽核效率提高，時效性將是待處理的問題。

1.2 研究目的

投保單位投保金額之稽核作業由受理登記、承保查核、案件處理、電腦查詢、電話溝通、實地訪查、罰鍰處分、行政救濟、強制執行等作業，其所花費的人力、物力成本所費不貲，如何有效的處理稽核作業以節省成本，能更有效率的完成使命，須借助於資訊科技所提供的有用資訊，幫助承辦人員於分析問題更能掌握問題核心，更能有效率的解決問題。本研究期望達到此一目的，我們希望藉由類神經網路與基因演算法的技術，以輔助稽核人員決策的進行與提升決策的品質。

2. 文獻探討

2.1 投保金額保費及負擔比率計算原則

全民健保的支出主要為醫療服務支出，其經費主要來源為保險費的收入，其中保險對象占38%、投保單位占37%、政府補助占25%。

為實現全民健保社會互助的理想，保險費的負擔比率也依保險對象類別不同而有差異，第一類公民營事業、機構有一定雇主之受雇者，被保險人負擔30%、投保單位負擔60%、政府負擔10%，二、三類則由被保險人與政府共同負擔保費，投保單位為免負擔，第四類（義務役軍人、軍校軍費生、替代役役男、在恤遺眷）、第五類低收入戶和第六類無職業的榮民，保險費完全由政府撥款補助[1]。

2.2 類神經網路

1943年,心理學家McCulloch 和數學家Pitts 首先提出一個神經元數學模式(MP模型),設計第一個數位式的神經元架構。在1949年Hebb 提出了第一個類神經網路學習法則,啟發後來學習法則方面的研究,1957年Rosenblatt 提出感知機模式,為最早的類神經網路模式,成功地用兩層感知神經網路(Perceptron Neuron Network)模擬人眼。經過了20多年,在麻省理工學院從事研究的Minsky和Poppert針對XOR問題提出質疑,批評神經網路的不適用性,使許多研究者轉而研究人工智慧,直到1982年美國加州理工學院物理學家Hopfield 提出霍普菲爾網路與倒傳遞網路,才解決許多研究上的瓶頸,使類神經網路應用於各個領域[2-11]。

類神經網路目前已有數十種不同的模式。我們依據研究問題之特性,選擇了監督式倒傳遞網路(Back-Propagation Networks, BPN)[12-17],針對投保單位投保金額的特性,加以分析。

2.3 倒傳遞網路

倒傳遞網路屬於一種前饋式網路架構,具有監督式學習的過程。自從Werbos於1974年提出加入隱藏層設計之網路學習演算法開始,再由Rumelhart和Hinton等人於1985年提出通用差距法則(general delta rule),並發表倒傳遞類神經網路理論後,其價值正式被肯定,發展至今已是目前最常被使用的類神經網路之一,因為此網路具有學習能力佳、回想速度快、輸出值可以為連續值等優點,能處理複雜的樣本識別與高度非線性的合成問題,因而廣泛應用於各個領域[5]。

Werbos和Parker雖已在1974年和1982年分別提出倒傳遞類網路基本概念。直到1985年,史丹福大學的Rumelhart、Hinton和Williams提出了反饋學習法則,或稱通用差距法則(Generalized delta learning rule),才將這項理論與演算法則清楚的定義。倒傳遞類神經網路其基本原理是利用最陡坡降法(Gradient descent method),將誤差函數最小化進而推導出誤差法則(Delta rule)。

2.4 基因演算法

基因演算法(Genetic algorithm)是一種求解最佳化的演算工具,其理論基礎乃源自於1859年達爾文的『物種演化』(On the Origin of Species by Means of Nature Selection)書中的『物競天擇、適者生存』的演化及淘汰觀念,

模擬生物之間的生存競爭,倖存者得以繁衍至下一代。然而,環境會隨著時間不斷改變,各種生物也會不斷的改變其特徵及能力,以適應不斷改變的環境,如此才能繼續生存。但有些生物的演化卻無法跟上環境的改變,因而產生生物間對環境適應力的高低差異,適應力高的生物其所留下的後代數目會較適應力低的生物多,如此一代一代演化下去,最後會演變進而產生一完全適應環境的族群,達到適者生存的概念[9]。

應用基因演算法搜尋最佳解的問題前,需要先將問題之決策變數進行編碼成稱之為染色體的數值型或是二元字串,接著計算適應函數的適應性是否為佳,如果適應性為佳也符合問題要求即為本問題的最佳解,反之適應性不佳則需要進行演化過程有複製、交配、突變等過程,當過程完成再次計算是否符合問題要求,不斷的重複隨機計算以上流程,直到搜尋符合問題之最佳解才停止。

基因演算法的演算流程是隨機產生初始族群,再經過選擇淘汰適應值較低的染色體,保留適應值較高的染色體,完成選擇與複製。然後再隨機選擇兩個篩選過的染色體,進行相互交配的動作,達成交配運算。經過交配過的族群,再依據突變率隨機性的選擇一些個體,改變染色體中的部分字串值,以完成突變。經過選擇、交配及突變後,就產生新一代的族群,如此不斷重複篩選、複製、交配及突變的運算,搜尋問題的最佳解,直到滿足終止條件為止。再將適應值最高的染色體經過解碼後,所得到的參數,即是此問題的最佳解,亦是完成整個基因演算法的搜尋過程[12,14]。而基因演算法基本上有下列三個特性:

- (1) 基因演算法的搜尋過程是利用所有可能產生的解進行編碼,而不是參數本身,所以沒有參數空間分析上的限制。
- (2) 基因演算法可以同時考慮問題空間上多個點而不是一個點,所以可以快速的獲得整體最佳解,同時可減少得到局部最佳解的機會。
- (3) 基因演算法只利用適應函數的資訊做擇優而不需要其他資訊作輔助,且機率規則方式去引導求解,而並非有明確的規則,所以方法簡單且強健,因此不會受不同系統或問題的種類而有所限制。

3. 實驗方法

3.1 方法架構

本研究主要以倒傳遞網路與基因演算法混合型求取倒傳遞網路之初值設定值，藉此找出倒傳遞網路較佳之網路模型。

本研究資料庫中有 18,736 筆資料，其中申報正常有 4,246 筆，低報有 13,247 筆，高報有 1,243 筆。投保單位相關的特徵值有行業別大類、類目屬性、規模、被保險人級距總值、單位眷屬人數總值、單位平均投保金額、單位性質、單位郵遞區號、勞保級距數差。本研究之流程如圖 1 所示。

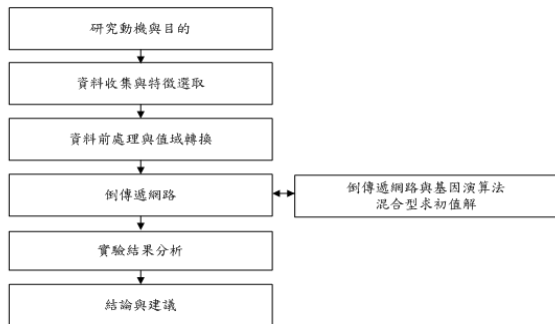


圖 1、研究流程圖。

3.2 資料前處理

本節介紹輸入變數之特徵值的分類、交叉分析表、值域轉換及二進制編碼。

3.2.1 特徵值分類

(1) 行業別:依行業系統表可分為農、林、漁、牧業、礦業及土石採取業、製造業、電力及燃氣供應業、用水供應及污染整治、營造業、批發及零售業、運輸及倉儲業、住宿及餐飲業、住宿及餐飲業、資訊及通訊傳播業、金融及保險業、不動產業、專業、科學及技術服務業、支援服務業、公共行政及國防、強制性社會安全、教育服務業、醫療保健及社會工作服務業、藝術、娛樂及休閒服務業、其他服務業等19項。

(2) 類目屬性:投保單位之類目屬性為中央非事業機關公教(職)人員、志願役軍人、中央以下機關學校公教(職)人員、地方公職人員、私立大專學校教職員、私立中小、私立高中、直轄市私立中小學校教職員、公營事業機構職員、公營事業機構工員、民營事業機構受雇者、非營利事業機構之受雇者、中央非事業機關、學校工員、私立學校工員、有一定雇主之受雇者、職業訓練機構之學員、公共行政、強制性社會安全等共有 15 項。

(3) 性質:投保單位之性質為行政、民意、司法、

文教、衛生、公營、軍事、一般、僑生、外籍生、海青班學員、獨資、合夥、公司等 14 項。

(4) 郵遞區號:投保單位含台北、基隆、宜蘭、金門、連江等縣市郵遞區號共有 72 項。

3.2.2 交叉分析表

投保單位輸入特徵值之申報狀態分析表。

(1) 行業別:表 1 為投保單位行業別大類之交叉分析表。

表 1、行業別大類交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
A	4	0.09	16	0.12	1	0.08	21	0.11
B	2	0.05	3	0.02	0	0.00	5	0.03
C	611	14.39	3560	26.87	312	25.10	4483	23.93
D	2	0.05	2	0.02	0	0.00	4	0.02
E	14	0.33	30	0.23	4	0.32	48	0.26
F	213	5.02	691	5.22	51	4.10	955	5.10
G	1530	36.03	5407	40.82	491	41.11	7428	39.65
H	12	0.28	55	0.42	2	0.16	69	0.37
I	44	1.04	111	0.84	17	1.37	172	0.92
J	140	3.30	407	3.07	41	4.34	588	3.14
K	129	3.04	506	3.82	31	2.49	666	3.55
L	3	0.07	9	0.07	3	0.24	15	0.08
M	1254	29.53	1337	10.09	172	13.84	2763	14.75
N	56	1.32	448	3.38	21	1.69	525	2.80
O	4	0.09	9	0.07	2	0.16	15	0.08
P	34	0.80	109	0.82	21	1.69	164	0.88
Q	104	2.45	212	1.60	24	3.14	340	1.81
R	11	0.26	21	0.16	3	0.24	35	0.19
S	79	1.86	314	2.37	47	5.71	440	2.35
	4246	100	13247	100	1243	100	18736	100

(2) 類目屬性:表 2 為投保單位類目屬性之交叉分析表。

表 2、類目屬性交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
11A	174	4.10	96	0.72	4	0.72	274	1.46
11B	326	7.68	56	0.42	6	0.42	388	2.07
11C	121	2.85	3560	26.87	194	26.87	3875	20.68
11D	28	0.66	2	0.02	2	0.02	32	0.17
11E	179	4.22	95	0.72	31	0.72	305	1.63
12F	91	2.14	691	5.22	178	5.22	960	5.12
12G	89	2.10	62	0.46	62	0.46	213	1.14
12H	2854	67.22	5407	40.81	318	40.81	8579	45.79
12M	289	6.81	319	2.41	51	2.41	659	3.52
13I	13	0.31	609	4.60	34	4.60	656	3.50
13J	9	0.21	506	3.82	36	3.82	551	2.94
13K	18	0.42	44	0.33	5	0.33	67	0.36

13L	31	0.73	1337	10.09	205	10.09	1573	8.40
13N	21	0.49	448	3.38	102	3.38	571	3.05
15P	3	0.07	15	0.11	15	0.11	33	0.18
	4246	100	13247	100	1243	100	18736	100

(3) 規模：表 3 為投保單位規模之交叉分析表。
表 3、規模交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
1-20	2547	59.99	10578	79.85	545	43.85	13670	72.96
21-40	859	20.23	897	6.77	207	16.65	1963	10.48
41-60	121	2.85	457	3.45	104	8.37	682	3.64
61-80	114	2.68	314	2.37	83	6.68	511	2.73
81-100	98	2.31	254	1.92	78	6.28	430	2.30
101-120	84	1.98	178	1.34	54	4.34	316	1.69
121-140	74	1.74	104	0.79	47	3.78	225	1.20
141-160	50	1.18	98	0.74	41	3.30	189	1.01
161-180	45	1.06	47	0.35	32	2.57	124	0.66
181-200	27	0.64	24	0.18	24	1.93	75	0.40
201-~	227	5.35	296	2.23	28	2.25	551	2.94
	4246	100	13247	100	1243	100	18736	100

(4) 被保險人級距總值：表 4 為投保單位被保險人級距總值之交叉分析表。
表 4、被保險人級距總值交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
1-100	511	12.03	1059	7.99	6	0.48	1576	8.41
101-200	39	0.92	906	6.84	58	4.67	1003	5.35
201-300	87	2.05	1874	14.15	41	3.30	2002	10.69
301-400	845	19.90	4132	31.19	174	14.00	5151	27.49
401-500	796	18.75	3149	23.77	14	1.13	3959	21.13
501-600	218	5.13	324	2.45	5	0.40	547	2.92
601-700	55	1.30	98	0.74	17	1.37	170	0.91
701-800	213	5.02	108	0.82	291	23.41	612	3.27
801-900	569	13.40	472	3.56	412	33.15	1453	7.76
900-1000	477	11.23	356	2.69	128	10.30	961	5.13
1001-~	436	10.27	769	5.81	97	7.80	1302	6.95
	4246	100	13247	100	1243	100	18736	100

(5) 眷屬人數總值：表 5 為投保單位眷屬人數總值之交叉分析表。
表 5、單位眷屬人數總值交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
1-500	1875	44.16	1059	7.99	19	1.53	2953	15.7
501-1000	853	20.09	906	6.84	62	4.99	1821	9.72
1001-1500	547	12.88	2318	17.50	37	2.98	2902	15.4
1501-2000	124	2.92	3679	27.77	180	14.48	3983	21.2
2001-2500	99	2.33	2187	16.51	24	1.93	2310	12.3

2501-3000	54	1.27	698	5.27	35	2.82	787	4.20
3001-3500	84	1.98	402	3.03	39	3.14	525	2.80
3501-4000	62	1.46	396	2.99	307	24.70	765	4.08
4001-4500	51	1.20	454	3.43	217	17.46	722	3.85
4501-5000	61	1.44	326	2.46	198	15.93	585	3.12
5001-~	436	10.27	822	6.21	125	10.06	1383	7.38
	4246	100	13247	100	1243	100	18736	100

(6) 平均投保金額：表 6 為投保單位平均投保金額之交叉分析表。
表 6、單位平均投保金額交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
1-10000	0	0.00	0	0.00	0	0.00	0	0.00
10001-20000	894	21.06	914	6.90	7	0.56	1815	9.69
20001-30000	1957	46.09	5107	38.55	492	39.58	7556	40.33
30001-40000	1067	25.13	4403	33.24	347	27.92	5817	31.05
40001-50000	210	4.95	2114	15.96	287	23.09	2611	13.94
50001-60000	107	2.52	632	4.77	107	8.61	846	4.52
60001-70000	5	0.12	21	0.16	1	0.08	27	0.14
70001-~	6	0.14	56	0.42	2	0.16	64	0.34
	4246	100	13247	100	1243	100	18736	100

(7) 性質：表 7 為投保單位性質之交叉分析表。
表 7、單位性質交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
1	58	1.37	547	4.13	9	0.72	614	3.28
2	24	0.57	5	0.04	1	0.08	30	0.16
3	36	0.85	24	0.18	2	0.16	62	0.33
4	114	2.68	1224	9.24	0	0.00	1338	7.14
5	98	2.31	621	4.69	0	0.00	719	3.84
6	84	1.98	178	1.34	5	0.40	267	1.43
7	18	0.42	16	0.12	3	0.24	37	0.20
8	1755	41.33	4755	35.89	536	43.12	7046	37.61
A	0	0.00	0	0.00	0	0.00	0	0.00
B	0	0.00	0	0.00	0	0.00	0	0.00
C	0	0.00	0	0.00	0	0.00	0	0.00
D	347	8.17	1647	12.43	294	23.65	2288	12.21
E	14	0.33	120	0.91	89	7.16	223	1.19
F	1698	39.99	4110	31.03	304	24.46	6112	32.62
	4246	100	13247	100	1243	100	18736	100

(8) 郵遞區號：表 8 為投保單位郵遞區號之交叉分析表。

表 8、單位郵遞區號交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
台北	2115	49.81	6119	46.19	501	40.31	8735	46.62
台北	1884	44.37	5017	37.87	406	32.66	7307	39.00
基隆	129	3.04	1127	8.51	312	25.10	1568	8.37
宜蘭	117	2.76	984	7.43	24	1.93	1125	6.00
金門	1	0.02	0	0.00	0	0.00	1	0.01
	4246	100	13247	100	1243	100	18736	100

(9) 勞保級距數差：表 9 為投保單位勞保級距數差之交叉分析表。

表 9、勞保級距數差交叉分析表。

	正常		低報		高報		總計	
	件數	百分率	件數	百分率	件數	百分率	件數	百分率
1-25	3419	80.52	3147	23.76	813	65.41	7379	39.38
26-50	421	9.92	956	7.22	174	14.00	1551	8.27
51-75	179	4.22	1874	14.15	71	5.71	2124	11.33
76-100	36	0.85	3151	23.79	97	7.80	3284	17.52
101-125	34	0.80	2174	16.41	34	2.74	2242	11.96
126-150	17	0.40	217	1.64	5	0.40	239	1.27
151-175	14	0.33	84	0.63	17	1.37	115	0.61
176-200	21	0.49	91	0.69	13	1.05	125	0.66
201-225	29	0.68	319	2.41	9	0.72	357	1.90
226-250	37	0.87	318	2.40	4	0.32	359	1.91
251-~	39	0.92	916	6.91	6	0.48	961	5.12
	4246	100	13247	100	1243	100	18736	100

3.2.3 值域轉換及二進制編碼

監督式學習的輸入處理單元雖然可以接受任意大小變數值，但是如果不同處理單元所接受的變數值域差別太大，將使小值域的變數之重要性無法表現出來，而由大值域的變數控制整個網路學習過程，因而影響學習效果。因此實際上將各輸入處理單元的值域統一，亦即變數值域轉換。變數值域轉換可分成三種，最大最小對映法、機率對映法、非線性對映法，本研究將使用最大最小對映法及非線性對映，最大最小對映法將應用於被保險人級距總值、單位眷屬人數總值、單位平均投保金額、單位郵遞區號、勞保級距數差等特徵值，此單位郵遞區號特徵值將經過投保單位低報由大到小排序後，在處理值域轉換。

非線性對映法將應用於規模特徵值，因規模此特徵值於 20 以下佔了 72.96%，亦即變數值在各區間的發生頻率極為懸殊，用最大最小

對映法、機率對映法二種線性對映法並不適當，所以在此應用非線性對映法。

二進制編碼應用於行業別大類、類目屬性、單位性質等特徵值，以二進制編碼如表 10~12 所示。

表 10、行業別大類二進制編碼表。

序號	行業別大類	二進制碼	序號	行業別大類	二進制碼
1	A	00000	11	K	01010
2	B	00001	12	L	01011
3	C	00010	13	M	01100
4	D	00011	14	N	01101
5	E	00100	15	O	01110
6	F	00101	16	P	01111
7	G	00110	17	Q	10000
8	H	00111	18	R	10001
9	I	01000	19	S	10010
10	J	01001			

表 11、類目屬性二進制編碼表。

序號	類目屬性	二進制碼	序號	類目屬性	二進制碼
1	11A	0000	11	13J	1010
2	11B	0001	12	13K	1011
3	11C	0010	13	13L	1100
4	11D	0011	14	13M	1101
5	11E	0100	15	15P	1110
6	12F	0101			
7	12G	0110			
8	12H	0111			
9	12M	1000			
10	13I	1001			

表 12、單位性質二進制編碼表。

序號	單位性質	二進制碼	序號	單位性質	二進制碼
1	1	0000	11	C	1010
2	2	0001	12	D	1011
3	3	0010	13	E	1100
4	4	0011	14	F	1101
5	5	0100			
6	6	0101			
7	7	0110			
8	8	0111			
9	A	1000			
10	B	1001			

倒傳遞網路輸出二進制編碼如表 13 所示。

表 13、申報狀況二進制編碼表。

序號	投保金額申報	二進制碼
1	正常	010

2	低報	100
3	高報	001

倒傳遞網路輸入與輸出配置如表 14 所示。

表 14、倒傳遞網路輸入與輸出配置表。

特徵值	二進制碼 (B) / 量化值 (Q)	輸入 (I) / 輸出 (O)	位元數	位元起迄
X1 行業別大類	B	I	5	Bit1-Bit5
X2 類目屬性	B	I	4	Bit6-Bit9
X3 規模	Q	I	1	Bit10
X4 被保險人級	Q	I	1	Bit11
X5 單位眷屬人	Q	I	1	Bit12
X6 單位平均投	Q	I	1	Bit13
X7 單位性質	B	I	4	Bit14-Bit17
X8 單位郵遞區	Q	I	1	Bit18
X9 勞保級距數	Q	I	1	Bit19
O1	B	O	3	Bit1-Bit3

3.3 混合式倒傳遞網路與基因演算法

傳統倒傳遞網路的初始參數往往影響其網路收斂的程度與速度，但其參數的設定並無一個明確的依循法則，較為常用的方法為試誤方式去尋找較佳的倒傳遞網路模型，本研究將應用基因演算法求解特性，將倒傳遞網路初值設定之隱藏層節點數、學習速率及慣性值透過基因演算法以求得倒傳遞網路較佳模型，如圖 2 所示。倒傳遞網路輸入層節點數為 19 個，本研究使用一層隱藏層，輸出層節點數為 3 個，最大疊代次數為 1000 次，最小誤差均方根為 0.001，其中隱藏層節點數、學習速率及慣性值等輸入值，將由基因演算法之染色體 5、6、6 位元基因組合而成，如圖 3 所示，隱藏層結節點為 bit1 至 bit5，學習速率為 bit6 至 bit11、慣性值為 bit12 至 bit17。

倒傳遞網路與基因演算法混合型演算法，須先設定族群大小、代數、交配率、突變率之初值，族群大小以 20、30、40、50 設定不同值以比較其差異，代數設定為 50，交配率設定為 0.8，突變率設定為 0.05，位元組越長，值域切割越細，得解精準度越高，但運算處理需花費時間成本較高，本研究因輸入及輸出共 22 個節點，所以隱藏層節點數其基因長度設為 5 個位元，值域為 1 至 31 個節點，每代中計算各組染色體之適應值及挑選(輪盤法)、交配(雙點交配)、突變等運作以完成各代的執行，其中計算各組染色體之適應值將以倒傳遞網路之

總對率表示。

倒傳遞網路輸入層節點數為特徵值編碼後共 19 個，輸出層節點數為申報狀態正常、低報、高報等共 3 個。

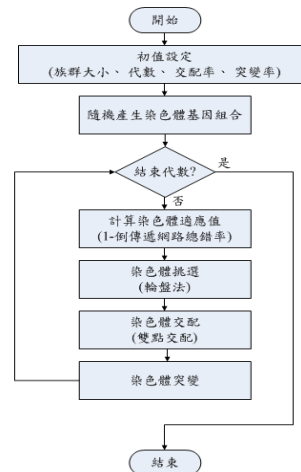


圖 2、倒傳遞網路與基因演算法混合型演算法。

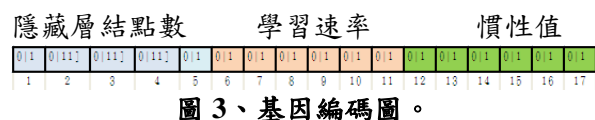


圖 3、基因編碼圖。

4. 實驗結果與分析

4.1 實驗目的

本研究期望建構出混合式倒傳遞網路與基因演算法最佳模型，對於投保單位投保金額低報能有良好的判別率，以精準率、召回率、F 測量指標分析其混合型效益以找出最佳模型。

4.2 模型分析

所提之混合式模型利用基因演算法求解的特性將每一世代中的族群大小算出適應函數值，再由輪盤法挑選出機率較大的適應函數值之染色體以複製到下一世代中，每一世代演化過程中將擁有較佳的適應函數值之染色體經由複製、交配、突變達到演化目的。

基因演算法交配率為 0.8，突變率為 0.05，不同族群大小 20、30、40、50 分別代表混合型一、混合型二、混合型三、混合型四，每代中所產生適應值的當代最大值、當代最小值、當代平均值的曲線圖，如圖 4~7 所示。

混合型一：族群大小=20。

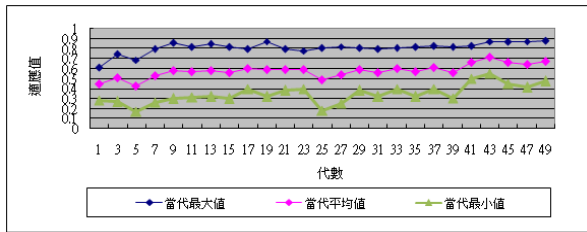


圖 4、混合型一之適應函數曲線圖。

混合型二：族群大小=30。

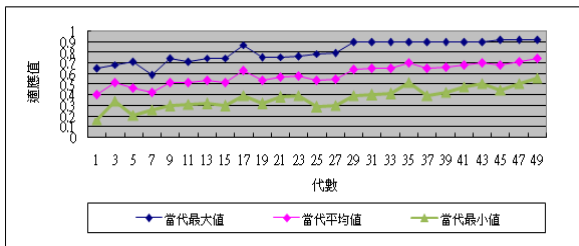


圖 5、混合型二之適應函數曲線圖。

混合型三：族群大小=40。

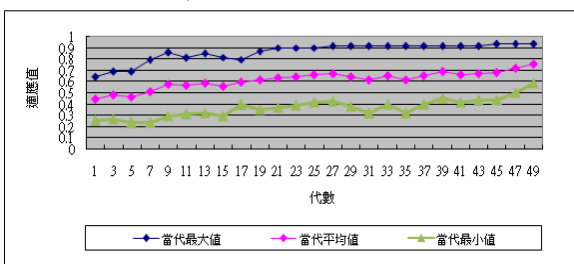


圖 6、混合型三之適應函數曲線圖。

混合型四：族群大小=50。

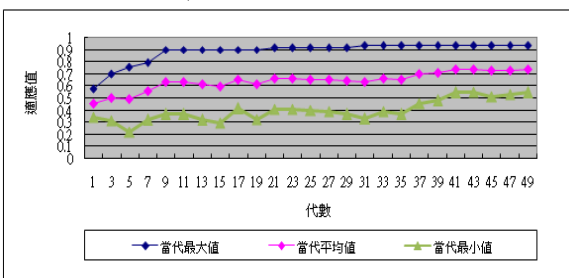


圖 7、混合型四之適應函數曲線圖。

4.3 模擬結果比較

本實驗四組模型分別為混合型一(隱藏層節點數=12、學習速率=0.23、慣性值=0.16)，混合型二(隱藏層節點數=11、學習速率=0.32、慣性值=0.21)，混合型三(隱藏層節點數=15、學習速率=0.14、慣性值=0.36)，混合型四(隱藏層節點數=10、學習速率=0.21、慣性值=0.25)，其成效評估以精準率、召回率、F 測量指標，如表 15 所示。結果分述如下。

表 15、混合型測試結果表。

	分類為低報單位	分類為非低報單位
低報單位	b	a
非低報單位	c	d
Precision=b/(b+c) Recall=b/(a+b) F=2PR/(P+R)		

混合型一：其隱藏層節點數=12、學習速率=0.23、慣性值=0.16、精準率=0.918、召回率=0.921、F1=0.919，如表 16 所示，

表 16、混合型一測試結果表。

	分類為低報單位	分類為非低報單位
低報單位	304(91.84%)	26(3.88%)
非低報單位	27(9.16%)	643(96.11%)
Precision=0.918 Recall=0.921 F1=0.919		

混合型二：其隱藏層節點數=11、學習速率=0.32、慣性值=0.21、精準率=0.899、召回率=0.869、F1=0.883，如表 17 所示。

表 17、混合型二測試結果表。

	分類為低報單位	分類為非低報單位
低報單位	287(89.96%)	43(6.31%)
非低報單位	32(10.04%)	638(93.68%)
Precision=0.899 Recall=0.869 F1=0.883		

混合型三：其隱藏層節點數=15、學習速率=0.14、慣性值=0.36、精準率=0.936、召回率=0.942、F1=0.938，如表 18 所示。

表 18、混合型三測試結果表。

	分類為低報單位	分類為非低報單位
低報單位	311(93.67%)	19(2.84%)
非低報單位	21(6.33%)	649(97.16%)
Precision=0.936 Recall=0.942 F1=0.938		

混合型四：其隱藏層節點數=10、學習速率=0.21、慣性值=0.25、精準率=0.891、召回率

=0.900、F1=0.895，如表 19 所示。

表 19、混合型四測試結果表。

	分類為低報單位	分類為非低報單位
低報單位	297(89.18%)	33(4.94%)
非低報單位	36(10.81%)	634(95.05%)
Precision=0.891 Recall=0.900 F1=0.895		

此實驗結果顯示最佳組合為混合型三其參數隱藏層節點數為 15、學習速率為 0.14、慣性值為 0.36，精準率為 0.936、召回率為 0.942，其 F 測量指標為 0.938。倒傳遞網路與基因演算法混合型網路進化產生四組解，但經測試資料測試四組不同網路架構，可發現於混合型一、混合型二、混合型四雖可達收斂，但對於低報之分類問題其精準率及召回率比較混合型三相對較低，由此可知混合型三，其對於誤判及漏判之分類較為敏感，故有最佳之 F 測量指標值。

5. 結論與建議

健保投保金額申報異常的問題，於不景氣的年代中更形嚴重，相關稽核的作業也較為複雜，不論其申覆往返的行政程序作業，在不擾民與稽核之間如何取得平衡點是一件困難的取捨，若以過往資料能找出最佳預測模型，能先過濾單位低報而先行處理，以提供單位與承辦人員能有更多資訊前置處理與應用，有效協助稽核人員掌握投保金額申報異常之投保單位，提升稽核人員之稽核效率，減少因稽核人員誤判造成投保單位必須執行申覆之煩瑣作業程序，本研究利用投保單位投保金額的相關屬性，經過資料前處理及值域轉換，建構出實用模型，以倒傳遞網路與基因演算法混合型為建構方法，實驗結果顯示本研究的預測效能可得正確辨識率達到 94.24%，比一般稽核人員判別的正确率 84.63% 高出許多。

倒傳遞網路雖已被普遍應用，但其初值設定並無明確的規則，往往只能以試誤方式去分析出最佳收斂之網路架構，倒傳遞網路與基因演算法混合型網路之進化求解，可將最佳之初值參數值求出，再利用精準率與召回率計算出 F 測量指標為其模式測量值，以求得混合型網路最佳解。

本研究利用不同族群大小為 20、30、40、50 共四種，測試其收斂狀況，其適應函數值皆

可收斂，此四種混合型再利用測試資料分析其精準率、召回率及 F 測量指標，當誤判筆數或漏判筆數減少時，F 測量指標值會增大，由 F 測量指標值中找出最佳倒傳遞網路模型。

由四種混合型中，最佳解為混合型三，其隱藏層節點數為 15、學習速率為 0.14、慣性值為 0.36，精準率為 0.936、召回率為 0.942、F 測量指標為 0.938。

本研究尚有一些問題可供後續相關研究者參考：

- (1) 投保單位的相關屬性，若有不同的屬性參考是否有不同的結果產生。
- (2) 倒傳遞網路與基因演算法混合型以初值設定為變數解，未考慮倒傳遞網路權重變數，若此演算法再加入此變數，對此效能上是否有所提升。
- (3) 應用不同演算法，例如 SOM 的模型，是否會產生不同的結果。

誌謝

本論文承國科會計畫編號 NSC97-2221-E-027-034-MY3 之補助得以順利完成，特此誌謝。

參考文獻

- [1] 中央健康保險局，
http://www.nhi.gov.tw/webdata/AttachFiles/Attach_5388_1_健全財務_970507_修.pdf
- [2] 王進德、蕭大全，*類神經網路與模糊控制理論入門*，全華科技圖書股份有限公司，台北市，2002 年。
- [3] 林文修，*演化式類神經網路為基底的企業危機診斷模型：智慧資本之運用*，國立中央大學資訊管理學系博士論文，桃園縣，2000 年。
- [4] 葉怡成，*類神經網路模式應用與實作*，儒林圖書有限公司，台北市，2004 年。
- [5] 葉怡成，*應用類神經網路*，儒林圖書有限公司，台北市，2001 年。
- [6] 張維哲，*人工神經網路*，全欣資訊圖書有限公司，台北市，1992 年。
- [7] 謝邦昌，*資料採礦與商業慧-SQL Server 2005*，鼎茂圖書出版股份有限公司，台北市，2005 年。
- [8] 蘇木春、張孝德，*機器學習、類神經網路、模糊系統以及基因演算法則*，全華科技圖書股份有限公司，台北市，1997 年。

- [9] 蔡瑞煌, *類神經網路概論*, 第七章, 三民書局, 台北市, 1995 年。
- [10] 范揚志, *應用類神經網路與基因演算法於射出成形製程參數最佳化之研究*, 中華大學科技管理學系碩士論文, 新竹市, 2005 年。
- [11] 周鵬程, *遺傳演算法原理與應用*, 全華科技圖書股份有限公司, 台北市, 2005 年。
- [12] Y.T. Wu, Y.J. An, J. Geller and Y.T. Wu, "A data mining based genetic algorithm," in *Proc. of The Fourth IEEE Workshop on Software Technologies for Future Embedded and Ubiquitous System*, Gyeongju, Korea, pp.55-62, Apr. 2006.
- [13] D. Casali, G. Costantini and M. Carota, "A cellular neural network based character recognition system," in *Proc. of IEEE Int. Symp. on Computational Intelligence in Robotics and Automation*, Espoo, Finland, pp. 481-486, June 2005.
- [14] Y. Liu, Z.G. Liu and J. Tang, "Hybrid method of BPN and genetic algorithm for completion time prediction," in *Proc. of Int. Conf. on Machine Learning and Cybernetics*, Guangzhou, China, pp.4625-4630, Aug. 2005.
- [15] J. Clark, I. Koprinska and J. Poon, "A neural network based approach to automated e-mail classification," in *Proc. of IEEE/WIC Int. Conf. on Web Intelligence*, pp.702-705, Oct. 2003
- [16] V. Chandiramani, R. Jayaseelan, V. Sudhan L. Nathan and K. Priya, "A neural network approach to process assignment in multiprocessor systems based on the execution time," in *Proc. of IEEE Int. Conf. on Intelligent Sensing and Information Processing*, Chennai, India, pp.332-335, Aug. 2004
- [17] I. Hayashi, M. Umamo, T. Maeda, A. Bastian, and L.C. Jain, "Acquisition of fuzzy knowledge by NN and GA- a survey of the fusion and union methods proposed in Japan," in *Proc. of the 2nd IEEE Int. Conf. on Knowledge-Based Intelligent Electronic Systems*, Piscataway, New Jersey, USA, pp.69-78, Apr. 1998.