

## 影響台灣 SARS 疫情擴散之因素分析 -以類神經網路及迴歸為預測模式

陸行<sup>1</sup>      金立人  
政治大學應用數學系

### 摘 要

2002 年十一月中國大陸廣東地區爆發嚴重急性呼吸道症候群(SARS) 疫情，疫情迅速漫延世界多國，本文以預防醫學角度提出十種可能影響 SARS 疫情擴散之因素：可能病例治癒人數、治療中之可能病例數、SARS 疑似病例數、死亡病例數、A 級居家隔離人數、香港感染人數、中國大陸感染人數、新加坡感染人數、越南感染人數及其他因素，並運用流行病學和醫學統計學等原理與方法，採用類神經網路與迴歸分析兩種研究模式來分析前九種因素對 SARS 新增感染人數的影響，給予定量分析評價，從兩種方法均可得到相同的結論。由本研究經類神經網路分析和迴歸分析驗證尚在治療中之可能病例數、死亡病例數、居家隔離人數是控制疫情的重要指標。尚在治療中之可能病例數會造成院內感染進而影響感染人數；死亡病例數的公佈能造成民眾更加小心而避免感染到 SARS；進行居家隔離能阻絕病源。所以，將來若有類似傳染病時，醫院應該作好防護設施，以及醫護人員應該要有良好的裝備，以避免院內感染的發生。並且應該立即制定良好的隔離政策，對於疫區來的人民、可能或疑似的病人及家屬實施居家隔離。此外公佈病患死亡人數，能使大眾更小心防範，對於控制疫情，也有不錯的效果。

關鍵詞：理論流行病學，倒傳遞網路法，迴歸相關係數。

---

<sup>1</sup> E-mail: slu@nccu.edu.tw



# A Study of Epidemic Outbreak of SARS in Taiwan by Neural Networks and Regression Models

Hsing Luh      Li-Jen Chin  
Department of Mathematical Sciences  
National Chengchi University

## Abstract

The outbreak of the deadly virus of severe acute respiratory syndrome (SARS) grew drastically in multiple ways of spreading from Southern China to many other countries in November 2002. This paper intends to study of the possible factors that relate to the number of SARS infections in Taiwan by Neural Networks and Regression Models. We investigate the number of new SARS cases by considering confirmed cases, suspected cases, probable cases, people with SARS-like symptoms, and number of infections in neighboring countries of Taiwan in that period. By two different study tools with confirmed analysis, we find that there are several important indices to track the number of new case occurrences, including probable and suspected cases, SARS infections and the number of people in quarantine, etc.

**Keywords:** Theoretical Epidemiology , Back-propagation Network , Regression



## 1、前言

2002 年十一月中國大陸廣東地區爆發嚴重急性呼吸道症候群(SARS) 疫情，疫情迅速漫延世界多國，台灣疫情也非常嚴重。世界衛生組織(WHO) 於 2003 年五月十九日對於台灣 SARS 疫情的控制有些評論，聲稱台灣為全球疫情擴散最快的地區，美國 CNN 電視台也報導由於未實施集中隔離，造成台灣疫情擴散快速。本文的目的在探討影響疫情擴散的因素，以作為未來若有類似傳染病時的參考依據。

SARS 是一種經由冠狀病毒引起，在人與人接觸中經由口腔、呼吸道感染的疾病，因為它是一種新的疾病，幾乎所有人對於它均無免疫力---皆可能感染 SARS。通常 SARS 會有二至七天的潛伏期，而後才發病，症狀初期為發燒、咳嗽、腹瀉、肌肉酸痛等，為期大約三天，此時為治療的黃金時期，接著大多數患者會退燒，此後病人之傳染力大增，接著一、二天後便再發燒，並且會有肺炎、肺部纖維化等症狀，大多數患者需要插管治療，可能引起呼吸衰竭而死亡。我們將所有的人歸納整理成四個族群：

- 1.易感染群(susceptible)：未感染 SARS 病毒的人。
- 2.潛伏群(incubational)：潛伏群則是已感染 SARS，但仍未有 SARS 症狀者。
- 3.發病群(ill)：已感染 SARS 病毒並已有 SARS 症狀者。
- 4.免疫或死亡群(immune or death)：免疫或死亡群則為曾患 SARS 症狀但已經痊癒，短時間內不會再受病毒感染或死亡。

本文以預防醫學角度提出十種可能影響 SARS 疫情擴散之因素，並運用流行病學和醫學統計學等原理和方法，分析前九種因素對 SARS 新增感染人數影響，給予定量分析評價：

- 1.可能病例治癒人數。
- 2.治療中之可能病例數。
- 3.SARS 疑似病例數。
- 4.死亡病例數。
- 5.A 級居家隔離人數。
- 6.香港感染人數。
- 7.中國大陸感染人數。
- 8.新加坡感染人數。



9.越南感染人數。

10.其他因素。

限於資料取得不易和療程至今仍未明朗，本文只針對前九項探討那些因素為影響感染人數的必要因素。研究工具是類神經網路及迴歸分析模式。

## 2、文獻探討

### 2.1 流行病學研究方法

預防醫學是現代醫學科學的重要組成部分，它是在以"預防為主"的衛生工作方針的指導下，根據人群中疾病的消長規律，採用環境衛生科學和基礎醫學等方法探查自然和社會環境因素對健康和疾病的作用規律，是研究人群中疾病發生、發展和預防的科學。運用流行病學和醫學統計學等原理和方法，分析環境中主要致病因素對人群健康的影響，給予定量分析評價，提出改善和利用環境因素的衛生要求和措施的理論。流行病學研究的方法則為了解傳染過程、傳染源、傳播途徑、人群易感性、疫源地及流行過程、影響流行過程的因素、傳染病的控制與消滅問題。而醫學統計的意義及基本概念包含研究因素、同質與變異、總體和樣本、參數和統計量、抽樣誤差、機率和機率推斷；統計資料類型包含計量資料、計數資料等；資料轉化統計工作步驟包含統計設計、搜集資料、整理資料、分析資料[1]。流行病學研究方法大致可分為如下類別。

#### (一)觀察法

1.描述性研究(descriptive study)

2.分析性研究(analytical study)

#### (二)實驗法

流行病學中所用的實驗法(experimental method)也叫作實驗流行病學(experimental epidemiology)，人群現場是流行病學的主要的、最大的實驗室。人群現場試驗(community field trial)對病因進行干預的又叫干預研究(intervention study)。

#### (三)理論和方法的研究

1.理論流行病學(theoretical epidemiology)研究也叫數理流行病學(mathematical epidemiology)研究，是將流行病學調查所得到的數據，建立



有關的數學模型(modeling)或用電腦模擬(computer simulation)，進行理論研究，又叫數理性研究(mathematical theory study)。

2.方法的研究 在著手一項特定研究之前，需要將研究中所使用的技術改善，發展收集數據資料的技術，改進疾病分類等[2]。

## 2.2 類神經網路與迴歸分析已廣泛應用於預防醫學研究

自 1997 年至今已有多篇論文採用類神經網路、迴歸分析作為各種醫學研究的模式。例如：Wong et al(2000)利用 Bayesian network 去學習香港醫院整形外科部門的醫療資料庫，分析各種因素的因果關係；廖上智、李憶農(2001)提出醫療資料遺漏值處理技術，運用統計及類神經網路兩種不同的數量分析方法測試，結果顯示此技術較簡單的取代法好；陳逸雯(2002)利用差值影像圈選可疑腫瘤區域，降低 False Positive 數目；羅佩禎、蘇祝鼎[4]以類神經網路來探討腦病變源的定位問題；陳永耀等(1998)利用類神經網路由溫度場反推換能器參數或血流量。若只使用迴歸分析模式，往往無法得到像類神經模式如此高的預測準確度。類神經網路與迴歸分析在實務研究與應用上一直扮演重要角色，自 1996 年至今已有多篇論文比較了兩者在不同領域之應用及其預測效果。大體而言，這些論文有以下共同的特點：

(1)均認為類神經網路的預測效果比迴歸分析準確。

(2)所採用的類神經網路都是使用前授型網路架構(feed forward structure)和監督型學習(supervised learning)。架構(structure)的建構大多採取試誤法(trial and error)，而學習(learning)方式則是使用倒傳遞(back propagation, BP)之方法(參考葉怡成(1997))。

雖然類神經分析法可以得到誤差度很小的預測結果，但是其最大缺點是無法求得數學關係式以及訓練的過程中無法得知需要多少神經元個數，太多或太少的神經元均會影響系統的準確性，因此往往需以試誤的方式得到適當的神經元個數。底下列出類神經網路優缺點：



### 類神經網路優缺點

優點：

- 1.可應用的領域相當廣泛。
- 2.可以接受不同種類的變數作為輸入，適應性強。
- 3.未知的輸入亦可得到正確的輸出，且有良好的推廣性。
- 4.可以建構非線性的模型。

缺點：

- 1.訓練的過程中無法得知需要多少神經元個數，太多或太少的神經元均會影響系統的準確性，因此往往需以試誤的方式得到適當的神經元個數。
- 2.解有無限多組，無法得知哪一組的解為最佳解。
- 3.以迭代方式更新鍵結值與閾值，計算量大，相當耗費電腦資源。
- 4.無法求得數學關係式(參考葉怡成(1997))。

因此，若單獨只用類神經網路模式或迴歸分析模式都各有缺點，所以本文採用類神經網路以及迴歸分析雙重模式彼此互相驗證，如此可以得到最佳預測分析效果。本文採用類神經網路為監督型學習(supervised learning)，架構(structure)的建構大多採取試誤法(trial and error)，而學習(learning)方式則是使用倒傳遞(back propagation, BP)之方法。

### 3、研究架構

由上述各國關於 SARS 的研究報告可使我們了解 SARS 疾病的各種知識，其中包括傳染過程、傳染源、傳播途徑、人群易感性、疫源地及流行過程，但是這些主要都是對於 SARS 病毒作描述性研究(descriptive study)以及分析性研究(analytical study)。目前比較缺乏理論流行病學方面的探討。本文中流行病學研究方法為理論流行病學(theoretical epidemiology)也叫數理流行病學(mathematical epidemiology)是將流行病學調查所得到的資料，使用電腦模擬(computer simulation)，建立有關的數學模型(modeling)及進行理論研究。

本文主要的研究目的是探討影響 SARS 流行過程的因素，因此在建構



數學模型(modeling)時，我們將 SARS 新增感染人數定為反應變數。因為台大使用靜脈免疫球(IVIG)來治療患者[8]，可提高患者存活機會，因此本文將可能病例治癒人數之因素列為自變數之一；由於 SARS 病毒在體外的存活性很強，而且由各國關於 SARS 的研究報告可知醫院環境，正好可使 SARS 病毒長期存活，因此本文將尚在治療中之可能病例人數、SARS 疑似病例之人數各列為自變數之一；此外我們對於一個爭議性的議題感到好奇，那就是媒體、政府該不該不斷的公佈病患死亡人數，到底公佈病患死亡人數對於 SARS 疫情的控制是否有幫助，或者只是會造成無謂的恐慌？因此本文將死亡病例人數列為自變數之一；並且由於國人普遍愛好自由，不太配合政府的隔離政策，因此本文將 A 級居家隔离人數列為自變數之一；此外，本文也考量鄰近疫區是否也會影響國內 SARS 新增感染人數。接著介紹資料的獲得。

### 3.1 資料背景時間

首先介紹資料收集，資料收集是自三月十七日 WHO 開始公佈各國疫情起，至五月十五日止，以日為單位，共六十日之實際資料。其中 50 個公佈日資料做為訓練樣本，最後 10 個公佈日資料做為預測比較之用，接下來探討 SARS 疫情的問題。

自 2001 年 11 月 SARS 疫情自中國大陸廣東爆發開來後，便在大陸各地開始漫延，但是由於大陸當局一直隱瞞疫情直到三月二十六日才公佈感染病例，使得臨近各國紛紛遭受感染。由圖 1 可知中國大陸 SARS 感染人數走勢圖呈現鋸齒狀分佈，這可能是由於中國大陸延誤通報所致。不過，由圖 1 可知中國大陸 SARS 感染人數已經過了最高峰，每日感染人數正在逐漸下降，自五月十三日起，每日通報病例也已經降到百名以下，這表示中國大陸 SARS 疫情已經獲得控制。

而由圖 2、3、4、5 可知中國大陸 SARS 疫情迅速傳染給香港、新加坡、越南、台灣等地區。越南疫情迅速獲得控制，自四月十四日發佈最後一名感染病例後，未再通報任何新增病例。越南已經從 SARS 感染地區除名。新加坡 SARS 感染人數走勢圖則呈穩定下降，自四月二十三日起，每日通報病例也已經降到五名以下，且自五月十一日至五月十七日無任何新增病例；而香港則經過幾波感染後，也正穩定下降，由走勢圖可知香港 SARS



感染人數的高峰已過，自五月六日起，每日通報病例已經降到五名以下。就感染人數而言，中國大陸、香港是屬於感染人數較多的地區，而新加坡、越南則是屬於感染人數較少的地區，這應該是與 SARS 發源地--廣東的距離及彼此間人民的互動往來有關；台灣地區 SARS 感染人數走勢圖呈現鋸齒狀上升。以下為台灣以及鄰近疫情較嚴重國家之感染人數走勢圖：

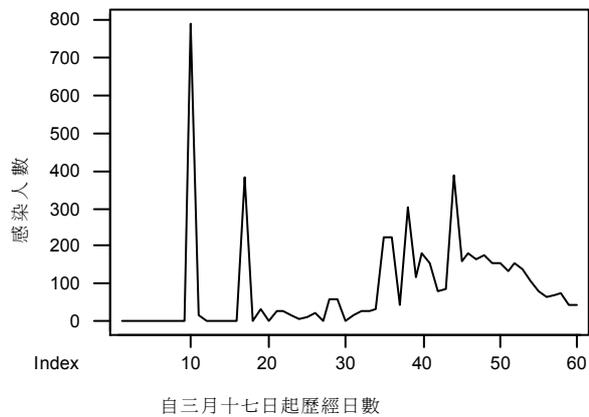


圖 1 中國大陸感染人數統計圖

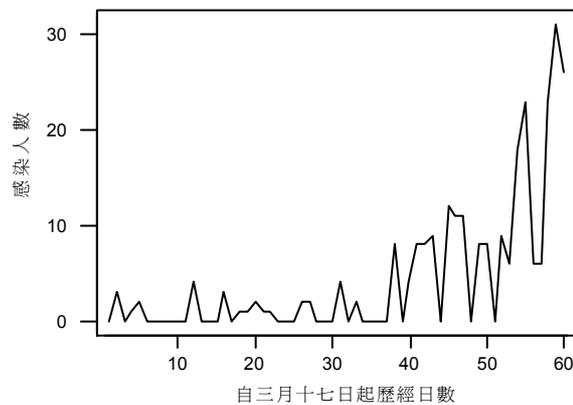


圖 2 香港感染人數統計圖

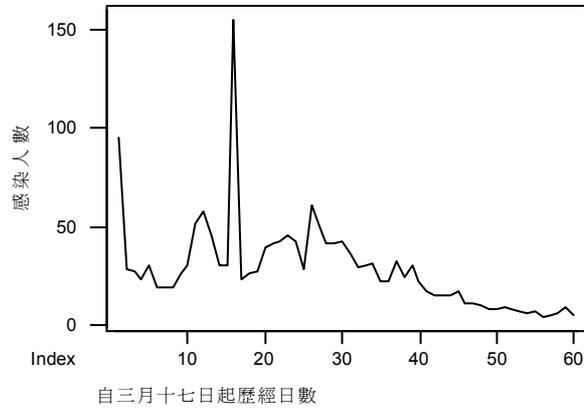


圖 3 台灣感染人數統計圖

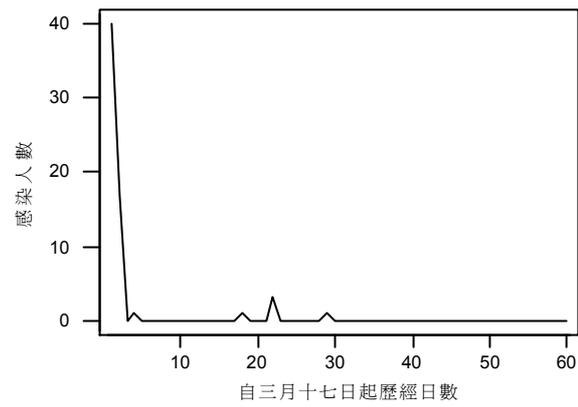


圖 4 新加坡感染人數統計圖

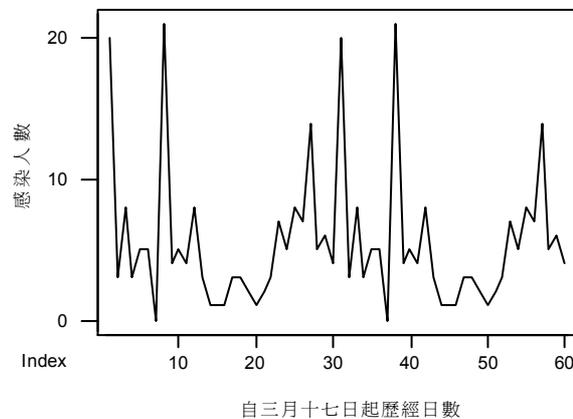


圖 5 越南感染人數統計圖

雖然這些國家位在亞洲地理區位上，各國之間存在真實的空間關係與直線距離，但是在探討各國之間人們經濟、旅遊的往來，實際地理區位關係無法完全解釋 SARS 疫情與環境互動的結果。本文探討東亞、東南亞國家 SARS 風暴對台灣的影響，以及台灣整體疫情、各國政府防疫措施才是影響 SARS 風暴擴散的主因。也因為這些國家的人民與台灣民眾常有密切的往來以及時空效應的擴散促使風暴如骨牌效應的展開；接下來介紹研究方法。

### 3.2 研究方法及模式

本文採用理論流行病學(theoretical epidemiology)研究法，針對九種可能影響 SARS 新增感染人數之因素：可能病例治癒人數之因素、尚在治療中之可能病例人數、SARS 疑似病例之人數、死亡病例人數、A 級居家隔离人數、香港感染人數、中國大陸感染人數、新加坡感染人數、越南感染人數作探討。採用類神經網路與迴歸分析兩種模式來分析。本研究的研究的流程是先將台灣以及鄰近疫區之疫情資料收集，如上節所述，以及確認資料無誤後，利用類神經網路分析各自變數對於預測反應變數的影響，再藉由迴歸分析建立迴歸模式。在類神經網路分析中，首先介紹基本設定；接著使類神經網路產生最佳預測，所得各項統計值；之後，列出各自變數在產生最佳預測時的重要性。在迴歸分析中，我們先進行線性迴歸效果顯著性檢驗，我們會得到相關的統計資料，以了解各自變數對於反應變數之

間是否存在線性關係；若有線性關係不顯著的自變數存在，則進行 Box-Cox 乘冪轉換，即對於預測反應變數無線性關係的自變數與反應變數之間的關係式進行轉換；接著再分析哪些自變數對於反應變數影響最大以及自變數彼此間是否有替代性而採用 Stepwise Regression。Stepwise Regression 包含三部份：Stepwise (forward and backward)、Forward selection、Backward elimination，限於篇幅只將 Stepwise (forward and backward)分析結果列出；之後，再建立 Best Subsets Regression 以評估最佳預測變數；最後再建立線性迴歸模式。圖 6 為本研究的流程表：

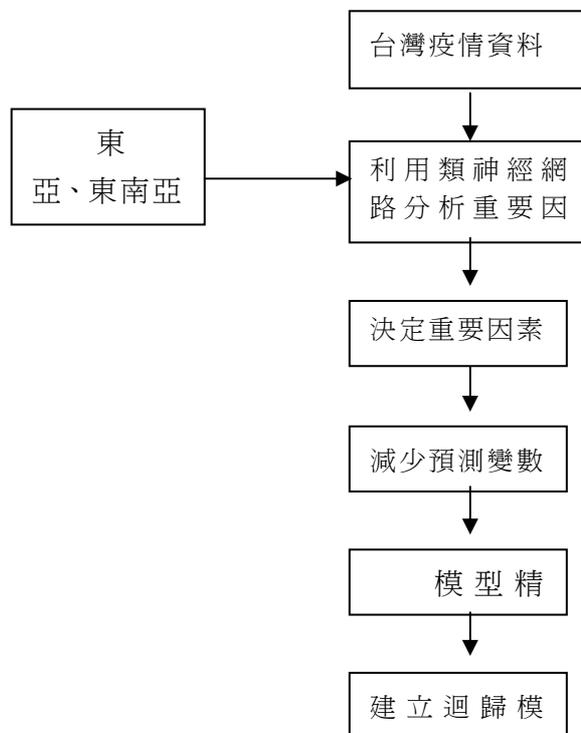


圖 6. 類神經網路及迴歸分析模式的建立流程

本文所採取的類神經網路為倒傳遞網路 (Back-propagation Network)，網路包含許多層，每一層包含若干個處理單元：

輸入層：以輸入自變數。

輸出層：以輸出反應變數。

隱藏層：提供類神經網路表現自變數間與反應變數間的交互作用，與問題

的內在結構的能力。以下顯示輸入與輸出的倒傳遞網路模型：圓圈的部分代表神經元底下為實際作業過程：使用分析軟體為 Neuro Shell 2(1993-1998)。

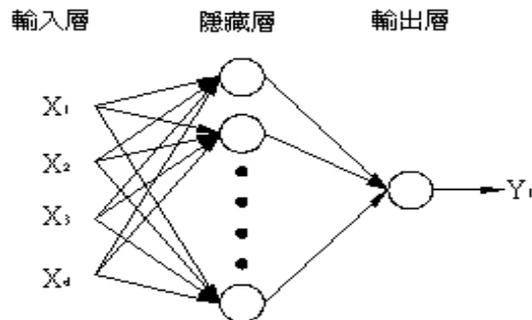


圖 7 倒傳遞網路模型[5]

本文所採用的隱藏層又可再細分成三層，見圖 7，各層的隱藏層節點、學習速率、動量係數設定採用試誤測試。而實驗的步驟流程為先將台灣疫情資料、東亞、東南亞疫情資料先藉由類神經網路分析，以決定重要因素。

#### 類神經網路模型：

自變數：

1. 可能病例治癒人數之因素  $X_1$ 。
2. 尚在治療中之可能病例人數  $X_2$ 。
3. 疑似病例之人數  $X_3$ 。
4. 死亡病例人數  $X_4$ 。
5. A 級居家隔離人數  $X_5$ 。
6. 香港感染人數  $X_6$ 。
7. 中國大陸感染人數  $X_7$ 。
8. 新加坡感染人數  $X_8$ 。
9. 越南感染人數  $X_9$ 。

反應變數：

SARS 新增之感染人數  $Y$

數學模型：

$$Y = F(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9) + \Phi_\varepsilon$$

其中  $Y$  為新增感染人數

其中  $\Phi_\varepsilon$  為誤差項，代表其他影響因素。

下列為利用類神經網路分析所得之各項統計值：

表 1 類神經網路分析所得之各項統計值

利用類神經網路分析所得之各項統計值	
R squared:	0.99950.
Mean squared error:	0.026
Mean absolute error:	0.069
Min. absolute error:	0
Max. absolute error:	0.831
Percent within 5%:	48.333
Percent within 5% to 10%:	0
Percent within 10% to 20%:	1.667
Percent within 20% to 30%:	3.333
Percent over 30%:	1.667

經過多次試誤後，我們能使類神經網路的預測值與實際值之間的 R squared 值相當高，Max. absolute error=0.831，且將近五成的預測值會落於實際值百分之五的誤差範圍中，表 1 顯示各項統計值。此時各層的隱藏層節點、學習速率、動量系數設定，如表 2 所示。並且由分析結果可知  $X_2$ 、 $X_4$ 、 $X_5$  為影響 SARS 新增感染人數  $Y$  的最重要因素，而  $X_7$ 、 $X_8$ 、 $X_9$  對於  $Y$  的影響最少，如圖 8 所示。但此時類神經網路的預測值卻也有將近五成會落於實際值百分之三十的誤差範圍外，因此必須採用迴歸模式以進一步分析確認。

表 2. 類神經網路模式的設定

INPUT	9
OUTPUT	1
隱藏層節點	4, 4, 4
學習速率	0.1, 0.1, 0.1
動量系數	0.1, 0.1, 0.1



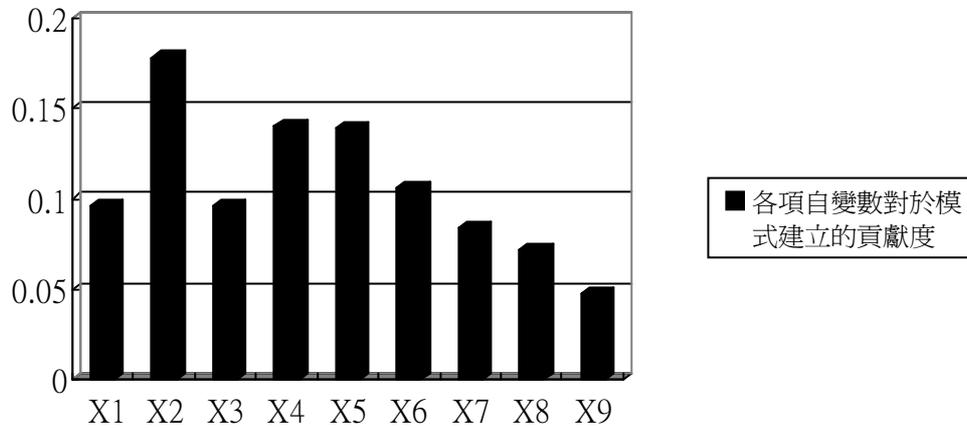


圖 8 類神經網路分析所得之各自變數對於模式建立的貢獻度

接著介紹迴歸模型設定：使用分析軟體為 MINITAB[12]。迴歸分析中將以  $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$  為研究變數，探討它們對於新增可能病例  $Y$  所造成的影響。

首先分別對  $X_1$  到  $X_9$  進行與  $Y$  的迴歸分析，以了解它們是否與  $Y$  有顯著的線性關係。令  $H_0: b_i = 0$ ， $b_i$  為迴歸模式中  $X_i$  的係數， $i=1, 2, \dots, 9$ 。定義  $F_\alpha(r_1, r_2)$  使得  $\Pr[F \geq F_\alpha(r_1, r_2)] = \alpha$  經由計算得： $F_{0.05}(1, 58) = 4.04$ ，由附錄簡單迴歸分析表可知  $F$  值小於 4.04 者只有  $X_7, X_8, X_9$ 。由此推論  $X_1, X_2, X_3, X_4, X_5, X_6$  對於預測  $Y$  值有較大的線性相關。藉由 Box-Cox 乘冪轉換測試[9]，我們發現它們對  $Y$  的非線性相關也不顯著。接著進行 Stepwise Regression，所得結果如下表所示：

表 3. Stepwise Regression

Stepwise Regression: y versus $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$ Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15 Response is y on 6 predictors, with N = 60				
Step		1	2	3
Constant		-0.2905	-0.2165	-1.9194
$X_2$	係數	0.128	0.193	0.355
	T-Value	11.98	7.90	5.85
	P-Value	0.000	0.000	0.000

$X_5$	係數		-0.00133	-0.00204
	T-Value		-2.91	-4.12
	P-Value		0.005	0.000
$X_4$	係數			-0.74
	T-Value			-2.89
	P-Value			0.005
	S	3.81	3.59	3.37
	R-Sq	71.22%	74.95%	78.20%
	R-Sq(adj)	70.72%	74.07%	77.03%
	C-p	16.3	8.9	2.8

Stepwise Regression 中可知  $X_2$  為影響  $Y$  值的最重要因素，除去  $X_2$  的影響後則  $X_5$  對於  $Y$  值的影響最大，並且除去  $X_2$ 、 $X_5$  的影響後則  $X_4$  對於  $Y$  值的影響最大，見表 3。再來進行 Best Subsets Regression 以評估最佳預測變數：

表 4 線型模式之統計值

**Best Subsets Regression: y versus x1, x2, x3, x4, x5, x6, x7, x8, x9**

Response is y

Vars	R-Sq (%)	R-Sq(adj) (%)	C-p	S	x x x x x x x x x														
					1	2	3	4	5	6	7	8	9						
1	71.2	70.7	14.6	3.8095		X													
1	64.8	64.2	30.4	4.2148					X										
2	74.9	74.1	7.4	3.5851		X			X										
2	73.8	72.8	10.3	3.6686		X	X												
3	78.2	77.0	1.5	3.3744		X	X	X											
3	76.8	75.6	4.9	3.4804		X	X	X											
4	78.6	77.0	2.5	3.3749		X	X	X											X
4	78.4	76.9	2.9	3.3864		X	X	X											X
5	78.8	76.8	4.1	3.3917		X	X	X	X										
5	78.7	76.8	4.1	3.3926		X	X	X	X										X
6	79.2	76.8	5.1	3.3908		X	X	X	X	X									X
6	79.1	76.8	5.2	3.3934		X	X	X	X	X									X
7	79.4	76.7	6.4	3.4020		X	X	X	X	X								X	X
7	79.3	76.6	6.6	3.4079		X	X	X	X	X								X	X
8	79.6	76.4	8.1	3.4217		X	X	X	X	X	X							X	X
8	79.5	76.3	8.3	3.4296		X	X	X	X	X	X	X						X	X
9	79.6	75.9	10.0	3.4539		X	X	X	X	X	X	X	X					X	X



由 Best Subsets Regression 評估最佳預測變數可知，只採用  $X_2$ 、 $X_4$ 、 $X_5$  為預測自變數能使 C-p 以及 s 值最小，並且 R squared=78.2% 和採用  $X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$ 、 $X_7$ 、 $X_8$ 、 $X_9$  為自變數時 R squared=79.6% 差不多，因此可知  $X_2$ 、 $X_4$ 、 $X_5$  是適當的一組預測自變數，見表 4。最後建立線性迴歸模式：

$$Y = -1.92 + 0.355 X_2 - 0.743 X_4 - 0.00204 X_5$$

表 5 線型模式之統計值

Predictor	Coef	SE Coef	T	P
Constant	-2.580	1.464	-1.76	0.084
$X_2$	0.35542	0.06073	5.85	0.000
$X_4$	-0.7430	0.2572	-2.89	0.005
$X_5$	-0.0020413	0.0004956	-4.12	0.000
s = 3.454                  R-Sq = 79.6%                  R-Sq(adj) = 75.9%				

由迴歸數學關係式可知： $X_2$  與  $Y$  為正相關，即  $Y$  因  $X_2$  的增加而增加；而  $X_4$ 、 $X_5$  與  $Y$  為負相關，即  $Y$  因  $X_4$ 、 $X_5$  的增加而減少。

#### 4、結果歸納分析：

藉由本文所進行理論流行病學 (theoretical epidemiology) 研究以及利用類神經網路分析，可進一步的了解尚在治療中之可能病例人數、死亡病例人數、A 級居家隔離人數是影響 SARS 新增感染人數之最重要因素，從迴歸分析亦可得到相同的結論。在迴歸分析中，發現中國大陸感染人數、新加坡感染人數、越南感染人數與 SARS 新增感染人數線性關係不顯著；接著在 Stepwise Regression 中可知尚在治療中之可能病例人數為影響 SARS 新增感染人數之最重要因素，並且由之後的 Best Subsets Regression 可知若只有尚在治療中之可能病例人數、死亡病例人數、居家隔離人數三個自變數與 SARS 新增感染人數之間的迴歸分析中 R-Sq 已達到 78.2% 與可能病例治癒人數之因素、尚在治療中之可能病例人數、疑似病例之人數、

死亡病例人數、A 級居家隔離人數、香港感染人數、中國大陸感染人數、新加坡感染人數、越南感染人數全為自變數時  $R\text{-Sq}=79.6\%$  相差不多；尚在治療中之可能病例人數、死亡病例人數、居家隔離人數是一組適當的預測自變數，因為可能病例治癒人數之因素、SARS 疑似病例之人數等自變數和尚在治療中之可能病例人數、死亡病例人數、A 級居家隔離人數有相當高的相關性。最後再建立迴歸方程式， $Y = -1.92 + 0.355 X_2 - 0.743 X_4 - 0.00204 X_5$ ，在迴歸方程式中只有  $X_2$  項的係數為正，而  $X_4$ 、 $X_5$  項的係數皆為負，這代表  $X_2$  值增加可造成  $Y$  值增加；但是  $X_4$ 、 $X_5$  增加卻造成  $Y$  值減少。其實際含意為尚在治療中之可能病例人數愈多會造成新增感染人數的增加；反之，死亡病例人數、A 級居家隔離人數愈多會造成新增感染人數的減少，其可能的理由為：尚在治療中之可能病例人數愈多，造成院內感染的機會多；反之，死亡病例人數愈多造成大家愈小心謹慎，因而能減少新增感染人數。

## 5、結論

瑞典病理學家 Folke Henschen 說過：「人類的歷史即是疾病的歷史」。致病病毒、細菌的基因常常因為突變或彼此不斷地交換改變，因此人類不斷地遭遇各種新疾病的侵襲。但是一種新疾病的疫苗、治療藥物的研發，往往需要很長的時間，在研發出疫苗、治療藥物之前，如何避免疫情擴散，就是當務之急。本文研究的目的，就是希望能找到影響台灣地區新型流行傳染病新增感染人數的因素，以期當有新型流行傳染病爆發時，能儘早對這些因素加以管控，而能控制疫情。

由本研究經類神經網路與迴歸分析兩種研究模式驗證尚在治療中之可能病例數、死亡病例數、居家隔離人數是控制疫情的重要指標。尚在治療中之可能病例數會造成院內感染進而影響感染人數；死亡病例數的公佈能造成民眾更加小心而避免感染到 SARS；進行居家隔離能阻絕病源。所以，將來若有類似傳染病時，醫院應該作好防護設施，以及醫護人員應該要有良好的裝備，以避免院內感染的發生。並且應該立即制定良好的隔離政策，對於疫區來的人民、可能或疑似的病人及家屬實施居家隔離。此外公佈病患死亡人數，能使大眾更小心防範，對於控制疫情，也有不錯的效果。



## 參考文獻

- 中國皖南醫學院預防醫學教研室，  
<http://www.37c.com.cn/literature/library/theory/012/01201003.html>，  
1999.
- 耿貫一(1995)，*流行病學，原理與方法*，第二版，第一卷，北京人民衛生出版社。
- 陳逸雯(2002)，*肺部腫瘤偵測之電腦輔助診斷系統*，中原大學醫學工程研究所碩士論文。
- 羅佩禎、蘇祝鼎(1999)，*類神經網路應用於腦病變源定位分析*，*中華醫學工程學刊* / 19:2，105-112。
- 陳永耀、謝銘鈞、嚴家鈺、陳啟鴻、林文澧(1998)，*類神經網路模型應用於超音波換能器與組織參數之估測*，*中華醫學工程期刊*，第 18 卷，第二期，129-138。
- 楊雅媛(2002)，*迴歸分析與類神經網路預測能力之比較*，政治大學統計所碩士論文。
- 葉怡成(1997)，*應用類神經網路*，儒林圖書公司。
- 陳定信(2003)，*科學人雜誌*，6 月號。
- 劉應興(1997)，*應用線性迴歸模型*，華泰書局。
- 廖上智 李憶農(2001)，*A Missing Data Treatment for Data Mining Applications in Medical Information Systems*，*高雄醫學科學雜誌*，198-206。
- Man Leung Wong, Wai Lam, Kwong Sak Leung, Po Shun Ngan, and J. C. Y. Cheng(2000), *Discovering knowledge from medical databases using evolutionary Magazine, IEEE Engineering in Medicine and Biology Algorithms*, Vol. 19, No.4, 45 –55.
- MINITAB*, 4.10.1998, 1990-1995 Microsoft Corp.
- NeuroShell 2*, Release 4.0, 1993-1998 Ward SystemsGroup Inc.



## 附 錄

### Regression Analysis: y versus x1

The regression equation is

$$y = - 1.12 + 0.402 x1$$

Predictor	Coef	SE Coef	T	P
Constant	-1.119	1.021	-1.10	0.277
x1	0.40190	0.05619	7.15	0.000

S = 5.176      R-Sq = 46.9%      R-Sq(adj) = 46.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1370.6	1370.6	51.16	0.000
Residual Error	58	1553.8	26.8		
Total	59	2924.4			

Unusual Observations

Obs	x1	y	Fit	SE Fit	Residual	St Resid
55	27.0	23.000	9.732	1.001	13.268	2.61R
58	30.0	23.000	10.938	1.132	12.062	2.39R
59	38.0	31.000	14.153	1.518	16.847	3.40R
60	41.0	26.000	15.358	1.671	10.642	2.17RX

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

### Regression Analysis: y versus x2

The regression equation is

$$y = - 0.290 + 0.128 x2$$

Predictor	Coef	SE Coef	T	P
Constant	-0.2905	0.6286	-0.46	0.646
x2	0.12821	0.01070	11.98	0.000

S = 3.809      R-Sq = 71.2%      R-Sq(adj) = 70.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2082.7	2082.7	143.52	0.000
Residual Error	58	841.7	14.5		
Total	59	2924.4			

Unusual Observations

Obs	x2	y	Fit	SE Fit	Residual	St Resid
48	67	0.000	8.300	0.590	-8.300	-2.21R
51	80	0.000	9.967	0.677	-9.967	-2.66R
56	125	6.000	15.736	1.066	-9.736	-2.66R
57	132	6.000	16.634	1.133	-10.634	-2.92R
58	152	23.000	19.198	1.330	3.802	1.07X
59	169	31.000	21.378	1.500	9.622	2.75RX
60	189	26.000	23.942	1.704	2.058	0.60X



**Regression Analysis: y versus x3**

The regression equation is

$$y = -0.227 + 0.0644 x3$$

Predictor	Coef	SE Coef	T	P
Constant	-0.2272	0.7999	-0.28	0.777
x3	0.064357	0.007340	8.77	0.000

S = 4.656      R-Sq = 57.0%      R-Sq(adj) = 56.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1666.8	1666.8	76.87	0.000
Residual Error	58	1257.6	21.7		
Total	59	2924.4			

Unusual Observations

Obs	x3	y	Fit	SE Fit	Residual	St Resid
51	191	0.000	12.065	1.061	-12.065	-2.66R
56	239	6.000	15.154	1.366	-9.154	-2.06R
57	252	6.000	15.991	1.452	-9.991	-2.26R
58	196	23.000	12.387	1.091	10.613	2.34R
59	265	31.000	16.827	1.540	14.173	3.23RX
60	281	26.000	17.857	1.648	8.143	1.87X

**Regression Analysis: y versus x4**

The regression equation is

$$y = 1.67 + 0.690 x4$$

Predictor	Coef	SE Coef	T	P
Constant	1.6747	0.6048	2.77	0.008
x4	0.68995	0.06682	10.33	0.000

S = 4.215      R-Sq = 64.8%      R-Sq(adj) = 64.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1894.0	1894.0	106.62	0.000
Residual Error	58	1030.4	17.8		
Total	59	2924.4			

Unusual Observations

Obs	x4	y	Fit	SE Fit	Residual	St Resid
45	1.0	12.000	2.365	0.579	9.635	2.31R
51	10.0	0.000	8.574	0.678	-8.574	-2.06R
55	18.0	23.000	14.094	1.085	8.906	2.19R
56	20.0	6.000	15.474	1.203	-9.474	-2.35R
57	24.0	6.000	18.233	1.446	-12.233	-3.09RX
58	25.0	23.000	18.923	1.508	4.077	1.04X
59	31.0	31.000	23.063	1.888	7.937	2.11RX
60	34.0	26.000	25.133	2.080	0.867	0.24X

**Regression Analysis: y versus x5**

The regression equation is

$$y = 0.800 + 0.00196 x5$$

Predictor	Coef	SE Coef	T	P
Constant	0.8002	0.8293	0.96	0.339
x5	0.0019605	0.0002705	7.25	0.000

S = 5.144      R-Sq = 47.5%      R-Sq(adj) = 46.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1389.7	1389.7	52.52	0.000
Residual Error	58	1534.7	26.5		
Total	59	2924.4			

Unusual Observations

Obs	x5	y	Fit	SE Fit	Residual	St Resid
48	4947	0.000	10.499	1.072	-10.499	-2.09R
51	6529	0.000	13.600	1.433	-13.600	-2.75R
59	6112	31.000	12.783	1.334	18.217	3.67R
60	5189	26.000	10.973	1.124	15.027	2.99R

**Regression Analysis: y versus x6**

The regression equation is

$$y = 7.64 - 0.114 x6$$

Predictor	Coef	SE Coef	T	P
Constant	7.640	1.330	5.75	0.000
x6	-0.11414	0.03613	-3.16	0.003

S = 6.559      R-Sq = 14.7%      R-Sq(adj) = 13.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	429.43	429.43	9.98	0.003
Residual Error	58	2494.97	43.02		
Total	59	2924.40			

Unusual Observations

Obs	x6	y	Fit	SE Fit	Residual	St Resid
1	95	0.000	-3.204	2.551	3.204	0.53 X
16	155	3.000	-10.053	4.652	13.053	2.82RX
55	7	23.000	6.841	1.146	16.159	2.50R
58	6	23.000	6.955	1.171	16.045	2.49R
59	9	31.000	6.612	1.099	24.388	3.77R
60	5	26.000	7.069	1.196	18.931	2.94R



**Regression Analysis: y versus x7**

The regression equation is

$$y = 4.24 + 0.00186 x7$$

Predictor	Coef	SE Coef	T	P
Constant	4.240	1.098	3.86	0.000
x7	0.001859	0.007023	0.26	0.792

S = 7.096      R-Sq = 0.1%      R-Sq(adj) = 0.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3.53	3.53	0.07	0.792
Residual Error	58	2920.87	50.36		
Total	59	2924.40			

Unusual Observations

Obs	x7	y	Fit	SE Fit	Residual	St Resid
10	792	0.000	5.712	5.042	-5.712	-1.14X
17	384	0.000	4.954	2.284	-4.954	-0.74X
44	389	0.000	4.963	2.316	-4.963	-0.74X
55	79	23.000	4.387	0.917	18.613	2.65R
58	73	23.000	4.376	0.921	18.624	2.65R
59	38	31.000	4.311	0.976	26.689	3.80R
60	39	26.000	4.313	0.974	21.687	3.09R

**Regression Analysis: y versus x8**

The regression equation is

$$y = 4.23 + 0.032 x8$$

Predictor	Coef	SE Coef	T	P
Constant	4.228	1.360	3.11	0.003
x8	0.0319	0.1861	0.17	0.865

S = 7.099      R-Sq = 0.1%      R-Sq(adj) = 0.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.48	1.48	0.03	0.865
Residual Error	58	2922.92	50.40		
Total	59	2924.40			

Unusual Observations

Obs	x8	y	Fit	SE Fit	Residual	St Resid
1	20.0	0.000	4.866	2.868	-4.866	-0.75X
8	21.0	0.000	4.898	3.045	-4.898	-0.76X
31	20.0	4.000	4.866	2.868	-0.866	-0.13X
38	21.0	8.000	4.898	3.045	3.102	0.48X
55	8.0	23.000	4.483	1.036	18.517	2.64R
58	5.0	23.000	4.387	0.919	18.613	2.64R
59	6.0	31.000	4.419	0.923	26.581	3.78R
60	4.0	26.000	4.355	0.953	21.645	3.08R

**Regression Analysis: y versus x9**

The regression equation is

$$y = 4.53 - 0.121 x9$$

Predictor	Coef	SE Coef	T	P
Constant	4.5266	0.9288	4.87	0.000
x9	-0.1206	0.1650	-0.73	0.468

S = 7.068      R-Sq = 0.9%      R-Sq(adj) = 0.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	26.67	26.67	0.53	0.468
Residual Error	58	2897.73	49.96		
Total	59	2924.40			

Unusual Observations

Obs	x9	y	Fit	SE Fit	Residual	St Resid
1	40.0	0.000	-0.296	6.492	0.296	0.11X
2	17.0	3.000	2.477	2.786	0.523	0.08X
55	0.0	23.000	4.527	0.929	18.473	2.64R
58	0.0	23.000	4.527	0.929	18.473	2.64R
59	0.0	31.000	4.527	0.929	26.473	3.78R
60	0.0	26.000	4.527	0.929	21.473	3.06R

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

