

利用資料萃取技術探索天氣、污染、健保記錄間關連性

韓欽銓

李修宇

中華大學資訊工程研究所

南華大學資訊管理研究所

cchan@chu.edu.tw

jason008@seeder.net.tw

(03) 5374281#8306

(05) 2721001#2013

摘要

自工業革命以來的百年間，由於科技的進步創造了無數的產品，雖然改善人們的物質生活，但也因工廠林立製造了大量的污染物質，讓我們賴以維生的地球遭到破壞，而有所謂之文明病的產生，人們因此嚐到自己所種下的苦果。本文嘗試利用資料萃取(Data Mining)的技術找出天氣、空氣污染、健保記錄間的關聯，提供學者做進一步的相關研究，有助於在科技發展與環境保護間找尋平衡點。本研究中將資料萃取技術應用在三個異質資料庫(天氣氣象(中央氣象局)、空氣污染(環保署)、健保(健保局))，這些資料彼此之間存在著微妙的關係，利用類神經網路中的BPN模型，建立起模擬預測的模型，藉以從中發掘知識。其方法是利用天氣氣象因子、空氣污染因子，並考量區域性的因子以加強預測的正確性，並藉著模型的建立來預測氣喘病的發作。最後，在本研究中建立一個大台北地區的氣喘病模型，我們發現模型中的預測值會與天氣、空污的反應間有延遲性的現象。這項結果顯示出現代人普遍不太重視自己的病症，大部份人不會在發病時立即就醫，會在病發後兩天才到醫院看病，才會造成時間落後(time lag)。

關鍵字：天氣氣象、空氣污染、資料萃取、氣喘、類神經網路、BPN

壹、前言

近百年來科技發達，創造了許許多多的產品改善人們的物質生活，但也同時製造了大量的污染物質讓我們賴以唯生的地球遭到破壞，人們因此得到一些以前從未聽聞或是提早面臨的病症，也就是一般人所稱的『文明病』。這些病症或許有藥可治或許是無藥可救，並且這些病的成因也讓專家覺得頭疼，因為這都是文明科技所帶來的副產品，是我們拋不開也丟不掉的。所幸，目前已有為數眾多的專家學者致力於污的防治及病症的醫療；可是，這些的成果或許不是五年十年就能有所成就的，所以

在他們的成果出現前，我們希望設計一套〔提供預警〕的機制，讓社會大眾使自己能趨吉避凶。

當前已有為數眾多的專家學者致力於污染防治及病症診治之研究；然而，這些研究並不是短期所能成就的，必須經過一連串『嘗試錯誤』(try and error)的過程，所以，希望在他們進行研究之前，能先為他們找尋一些蛛絲馬跡，減少走錯路的時間，更希望本研究所建立之模型能夠提供一些方向，作為專家找出疾病的因果關係甚至是治療方法的指引。

本研究中僅以大台北地區的氣喘病症發作模型為應用實例代表，將來會朝向建立起全

台各類相關疾病模型而努力。

道發炎，導致氣喘的發作。

貳、文獻探討

● 空氣污染之對於氣喘病

過去曾有研究[3]，對高屏地區 11-16 歲國中學生作調查，找尋室內、外環境因素對於氣喘發生之關連性，發現在室外的環境因素與氣喘的相關性中以空氣污染最為顯著[4]。而進一步校正干擾因子後，發現空氣污染中又以懸浮微粒、二氧化氮、一氧化碳、臭氧與氣喘好發之相關程度最高。並且在 Warlaw [27]的研究中也指出：

1. 臭氧及二氧化硫可能會造成氣管收縮，並刺激呼吸道做直接的反應。
2. 吸入低濃度的污染物會引起呼吸道發炎，而高濃度的污染物則是直接對呼吸道上皮造成毒性的作用。
3. 空氣污染會改變免疫系統的反應，因此易引起氣喘的過敏反應。

● 氣象因素之對於氣喘病

然而，引起氣喘病發作因素不僅於天氣污染，還有氣象的因素...等等，以下列舉幾項顯示氣喘與天氣變化有關之佐證[4,19]：

1. 春夏季傍晚間氣喘會加重，這是因為與病人吸入花粉有關。因為空氣中的花粉經過白天的曝曬後隨風飄散，到了晚間因氣溫下降使得氣流向下，花粉也隨之下沉，故而此時花粉在空氣中的濃度升高，所以病人發病的機率大增。
2. 秋冬季天氣變化無常，病人突然吸入冷空氣，會讓呼吸道急遽收縮，引起氣喘症狀。
3. 夏季溫度高，加上相對濕度大，使得細菌容易滋生，一旦吸入後容易引起呼吸

而空氣的污染也讓我們居住的環境遭受到破壞，比如：酸雨[8,12,13]是由二氧化硫及二氧化氮所造成；二氧化碳[7,16]則是造成溫室效應的主要原因，進而使得全球都遭到熱浪的襲擊，並且因為氣候改變造成沙漠化的形成[20]；再則是失去臭氧層[15]的保護，將使地球生物圈暴露在更多的輻射線下，可能會造成：人類皮膚癌、白內障等疾病罹患率的增加。經由以上的說明，可以知道天氣、空氣污染對氣喘病的發作存在相依性，於是便可以藉著天氣氣候、空氣污染與健保的記錄提供做預測分析，了解相關要素的變化。[1,5]

● 資料萃取技術

『資料萃取』(Data mining)是近年來資料庫應用領域中，相當熱門的議題。它的起源可追溯到二次世界大戰前，美國政府就已經在人口普查以及軍事方面使用這項技術。此技術著重於資料庫建立、資料分析與預測等相關的資訊科技技術，其目標則是為了發掘出新的、未知的規則或樣式[28,33,35,36,37]。它是一項利用電腦快速計算的能力，可以從大量的資料中不斷探索，以發掘隱藏在繁雜資料中的知識與規則的過程。以下是本研究整理各學者對資料萃取的定義。

<表一：學者對 Data Mining 的定義>

姓名	對 data mining 所下的定義	出處
朱慧德 周卓定	從資料中發掘資訊或知識。利用資訊科技的輔助分析資料以得到有用的資訊，並建立模擬真實世界的模型，並利用模型描繪出資料的特徵與關係。	[6]
林文修	從資料中發掘有用的資訊及知識(有人稱為 Knowledge Discovery in Database, KDD)，也有人稱為『資料考古學』	[14]

Cabena Hadjini an Stadler Verhees Zamasi	將先前是未知的，有效的資訊從龐大的資料庫中萃取出來的程序，並且將萃取出的資訊提供給主官(管)做為決策的輔助。	[23]
Linoff	面對大量的資料，使用自動或半自動的方法進行分析動作，藉以發掘有意義性的關係或是法則	[24]
Grupe Owran g	對於現在有的資料找出新的事實及發掘出專家尚未發覺的新關聯性	[24]
Yongji an	在大型資料庫中，辯視出有意義的樣式之過程	[24]
Fayyad Smyth	屬於 KDD 的一個過程，可以用於某些計算技術上，在容許的運算效率限制下，產生些可以建立於資料之上的特定樣式。	[24]
林信忠	其主體由資料(庫)、相關技術所構成的，目標在於發掘新的、未知的法則或樣式	[11]
張瑋倫	將資料庫中看似無用的資料轉化成有用的資訊，並找出隱藏的關聯法則。	[28]
李紹倫 陳富中	利用分類、關聯性、群聚分析、序列分析、行走路徑分析、機器學習及統計方法，從雜亂且資料量龐大的資料庫中找出隱藏的、未知的、卻又對企業經營非常有用的資訊。	[29]
Akeel Al-Att ar	可以讓複雜的企業程序成為易瞭解且可重新設計，並發掘出過去營運模式中一些有意義的樣式，而藉著發展這些模式就可以用來改善經營績效及避免企業經營危機。	[8]
郭展盛	需要特別的演算法輔助，才能抽取出特定型式資料集中有意義的樣式或規則。	[31]
Chung Gray Mannin o	認為知識發掘的重點為在資料中發掘出有用的知識，而資料萃取主要著重在一些為了從資料截取知識所用的演算法應用上面，因此知識發掘較具學術研究的意味，資料萃取則為其操作過程。	[34]

綜合前述學者對資料萃取的定義可知，「資料萃取」主體由資料或資料庫、相關技術(偏重於資訊科技)所構成，其目標則是為了發掘出新的，未知的規則或樣式。這些樣式有兩種用處，第一，瞭解資料的特徵與關係可以提供做決策時所需要的資訊。第二，資料的特徵可以幫助做預測。

在[11]將資料萃取分為「資料選擇」、「資料前處理與轉換」、「資料萃取」、「詮釋/評估」這四個階段。

1. 資料選擇:主要目的在於建立目標資料集(target data set),經過考慮後決定需要萃取何種資料?
2. 資料前處理與轉換:資料萃取出來後,對於不適用的資料(比如:重覆或null)預先作處理,再將資料轉換成工具能讀取的資料格式。
3. 資料萃取:藉由資料萃取軟體找出隱藏在資料背後的規則、特性、樣式。
4. 詮釋/評估:分析與解釋實驗中產生的結果。

並且「資料萃取」這階段可以建立下列六種模式[29] [32] [33] :

1. **Classification**: 是根據一些變數的數值做計算,再依照結果作分類。常被用來處理對象篩選的問題,我們會用一些已經分類的資料來研究它們的特徵,然後再根據這些特徵對其他未經分類或是新的資料做預測,這些用來尋找特徵的已分類資料可能是來自現有的歷史性資料,或是將一個完整資料庫做部份取樣,再經由實際的運作來測試。
2. **Regression**: 是使用一系列的現有數值經過統計與分析後,來預測一個連續數值的可能值。
3. **Time-Series**: Time-Series 與 Regression 很像,不同點在於它所分析的數值都與時間有關,利用現有的數值來預測未來的數值。Time-Series 的工具可以處理有關時間的一些特性,例如時間的階層性、季節性、節日、以及其他的一些特別因素如過去與未來的關連性有多少。
4. **Clustering**: 將資料分為幾組,其目的是要將組與組之間的差異找出來,同時也要將一個組之中的成員的相似性找出來。
5. **Association**: 是要找出在某一事件或是

資料中會同時出現的東西，也就是如果 A 是某一事件的一部份，則 B 也出現在該事件中的機率有 X %。

6. **Sequence**：與 Association 很類似，所不同的是 Sequence 中有關連的資料是以時間來區分先後關係，例如：如果做了 X 手術，則 Y 病菌在手術後感染的機率是 45%。

在[14]中提到Glymour等人的研究，建議資料萃取過程的進行步驟：

1. 理解資料與進行的工作；
2. 獲取相關知識與技術 (Acquisition)；
3. 融合與查核資料 (Integration and checking)；
4. 去除錯誤或不一致的資料 (Data cleaning)；
5. 發展模式與假設 (Model and hypothesis development)；
6. 實際資料萃取工作；
7. 測試與檢核所挖掘的資料 (Testing and verification)；
8. 解釋與使用資料 (Interpretation and use)。

因為近年來資訊科技的發達，使得資料萃取所需要計算能力得到滿足，並結合類神經網路的機器學習，來進行特徵的抽取及趨勢的預測。

以往類神經網路在各領域有著許多的應用，特別是在機器學習及預測[21,22]上有著顯著的成效。本研究中使用類神經網路中的BPN (Back-propagation Neural Network) [9,18]，是一種具有學習能力的多層式前授網路，該網路是由Rumelhart、McClelland 在1985年所創立的。BPN的運作分成“學習”與“回想”兩個部份；學習過程中是以監督式學習的方式，將問題領域中得到訓練樣式及目標輸出，再將其導路網路中，反覆調整權重值；回想過程是分類跟預測過程，可以讓網路告訴我們最可能的輸出結果[2]。利用資料萃取技術在大型資料庫中定義其演算法與分類規則的項目[25,26]，並且以類神經網路的探勘技術，監督式學習的特性，提供輸入/出的相關特徵因子，在動態或相互抑制的神經網路下，得到有相關的神經元。之後在加以預測，得到快速一套預測的模

型系統。

然而，我們需要從這三個維度中，抽取出真正會產生影響的因素，才能確保預測模型的準確率，所以將對各維度的主要因素及影響做個分析：

一、天氣氣象的因素：

由[7]可知天氣氣象原始資料的欄位有：

測站號碼 時間 氣壓 氣溫 濕度 露點
雲狀 雲量 風 降水 日照時數 能見度 天氣
狀況 地面狀況 天氣及視線 地中溫度 降水
紀錄 錯誤記錄。

在以往的研究報告中，發現幾個氣候與空污間的相關訊息[8,10,13]：

1. 在同時考慮溫室氣體的增溫作用與氣溶膠冷卻作用的情況下，所有的氣候模式皆推估出一台灣鄰近地區的平均氣溫將持續上升。在二氧化碳增為1.9倍時，溫度將上升0.8-2.4°C，四季增溫的程度無差異。
2. 溫室氣體增加的同時，大氣的懸浮微粒也會相對地增加，因為這樣的緣故也造成冷卻的作用。
3. 降水量的增加有效降低揚塵，可使得懸浮微粒大量地減少。

再則，與台大大氣科學所專業觀測員Eric Ma 協詢下，得到幾個影響的因素：

1. 氣壓：若該地區有高壓盤旋使空氣中污染物停滯，造成污染的加劇。
2. 空氣中的懸浮微粒可以幫助霧氣凝結，形成雨雲。
3. 飽和濕空氣中的絕對濕度是隨著溫度而有顯著變化的。溫度每上升攝氏10度，絕對濕度約可以增加近一倍[16]。

4.日照量充足將有效地幫助光化學反應，繼而產生光化學煙霧[1]，對人體健康產生危害。

所以，本研究經初步決策分析後，定義出天氣氣候中會對氣喘及空污的主要影響因素如下：

氣壓 氣溫 溫差 相對濕度 風速 日照率
降水量

二、空氣污染的因素

空氣污染物分為一次污染物及二次污染物：

一次污染物：是指直接從污染源排放的污染物質，如：二氧化硫、一氧化氮、一氧化碳、懸浮微粒等。

二次污染物：是指一次污染物在大氣中互相作用，經過化學反應或光化學反應所生成的，其毒性比一次污染物還強。常見的，如：硫酸及硫酸鹽氣溶膠、硝酸及硝酸鹽氣溶膠、臭氧以及活性中間物(又稱自由基)—HO₂、HO等。由環保署[7]所提供的7項污染物指標中，抽取出5項：一氧化碳 臭氧 氮氧化物 二氧化硫 懸浮微粒。

三、健保記錄

在國衛院的健保資料庫中申請到的”門診處方及治療明細檔”裡面一共有37個欄位，從中抽取出與本研究有關的欄位，分別是：醫事機構代碼(跟醫事機構基本檔比對後，可以找出該醫事機構為於在哪個縣市)就醫日期 國際疾病分類號。

參、資料萃取與預測模型的建立

在研究中，資料庫的建立：

因為牽涉三個維度的資料庫，候選的屬性會產生連鎖的反應，故先選取幾個文獻中所

提到的指標做研究，先找出整體的趨勢後再加入其它的屬性作微調的效果。

<表二：三維度資料庫所選欄位>

維度	所選取的屬性
氣象資料庫	平均氣壓、平均氣溫、溫差、相對濕度、平均風速、降水量、日照率
空氣品質資料庫	一氧化碳、臭氧、二氧化氮、二氧化硫、懸浮微粒
健保資料庫	就醫科別、就醫日期、出生日期、醫事機構代號、性別、國際疾病分類號

整個預測系統的開發原則是天氣及空污的條件來預測氣喘病的發作。其建置步驟如下：

1. 氣象資料的收集：向中央氣象局購買86年度全台的逐日記錄後，並對其做前處理(含無效值及缺值插補)。
2. 空污記錄的收集：在環保署的網頁上下載86年度全台的逐日空氣污染記錄報告，並對其做前處理(含無效值及缺值插補)。
3. 健保記錄的收集：向國衛院申請的86年度”門診處方及治療明細檔”，先挑選出是”氣喘病”的病例記錄，並與醫事機構檔對應，以找出病人看診的地區，並對其做前處理(含無效值及缺值插補)。
4. 資料的整合及轉換：解讀資料欄位及轉換資料後，選擇我們所選擇的必要欄位匯進資料庫。
5. 知識探索的階段：因為Matlab對數學運算的能力很強，將有助於處理效率的提昇。所以挑選Matlab作為撰寫BPN的語言開發環境。將先前的三異質資料庫一起轉進Matlab，使其匯整成同一個表單，以降低資料庫轉換的時間，避免類神經網路學習中發生錯誤的可能。在類神經網路訓練出

來之後，將測試資料及未知性的資料放進模型中，以驗證訓練成果。

本研究中以大台北地區(台北縣市)的氣喘病人作為研究對象，這樣的作法可以消除地域因素，以避免地理條件的不一致。研究中的訓練樣本是 86 年度台北縣/市各站點的實際記錄，而各站點的每日記錄平均值作為研究中的測試樣本。本研究中所定義的”氣喘發病機率 Prob.”如下：

Prob.=當地每日氣喘病看診人次/年度當地氣喘病看診的人次

肆、實驗結果與討論

在本研究中建立一個大台北地區的氣喘病預測模型，嘗試利用類神經網路與資料萃取的技術找出天氣、空氣污染、健保記錄間的關聯。

首先在氣象測站方面，大台北地區一共有 4 個站點，去除鞍部站後尚有淡水、台北、竹子湖等 3 站。我們嘗試以氣象站與空污站配對的方式來找出北市空污站與北縣空污站應該配哪個氣象站，才會讓模型的預測準確率提高。圖 1、2、3，是以北市的空污站配上氣象台北及竹子湖的實驗結果，可以看出北市空污配上氣象台北的準確率比較高。

再則以北縣的空污與氣象台北及淡水作配對，實驗的結果顯示出北縣空污配上氣象淡水準確率會比較高(見圖 4、5、6)。

之後將北市空污+氣象台北與北縣空污+氣象淡水合併起來成為大台北地區，並把氣喘病機率做重新計算(因為樣本空間不一樣了)。而大台北地區所訓練的成果如圖 7 所示。

然而這樣的成果並不足以讓我們滿足，於是我們想到或許是病人延遲就醫的緣故，使得預測的結果產生位移的效果，所以嘗試做了包含次日記錄(圖 8)、後二日記錄(圖 9)、後三日

記錄(圖 10)的模型，結果在後二日的部份準確率比較高，所以我們找出大台北地區的病患傾向會在發作兩天內前往就醫。

伍、研究限制與問題解釋

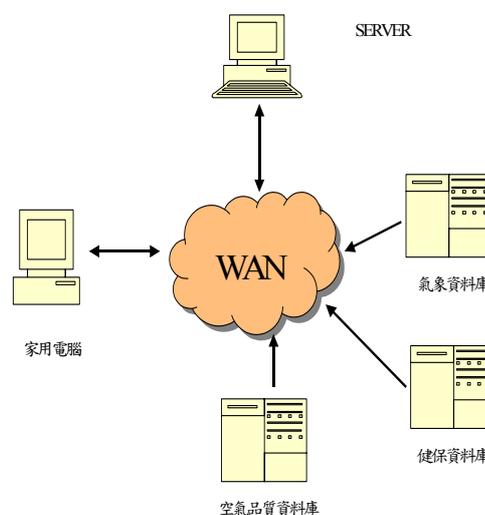
由於，類神經網路先天性就無法加入’時序性’因素，所以研究中加入延後幾日的考量，但這樣的作法並未能忠實地將原始的狀態還原，故產生預測上的誤差。希望以後研究改採用隱藏式馬可夫模型(HMM)或其它能考慮時序性的模型。

天氣與空污的記錄中，自動間測儀器的失靈及無效值，使得很多記錄產生錯誤，而研究中僅有 1 年的記錄樣本，所以又不能把那些記錄消除，因此僅能以平均值回填，這也是誤差產生的原因。

由於環保署及氣象局未能提供逐時記錄，再加上病人就診的時間與發發的時間又有延遲，所以模型的敏銳性不高，僅能得到概括性的效果。

陸、未來研究方向

本研究提供一個預警機制的建立，但因為時間及系統功能整合的問題，無法立即呈現，待日後後續研究時繼續努力。



<圖一：預警機制架構圖>

1. 使用者透過家裡的電腦上網，到達我們所設置的網站。在網頁上下達”查詢”的指令系統會將其送達遠端的 Server。
2. Server 收到”查詢”的指令後，下達 Query 的指令給氣象的資料庫，等待回應。
3. 氣象資料庫回應後將最新的氣象記錄傳送給我們的 Server。
4. Server 再下達 Query 給空氣品質資料庫，等待回應。
5. 空氣品質資料庫回應後將最新的空氣品質記錄傳送給我們的 Server。
6. Server 再下達 Query 給健保記錄資料庫，等待回應。
7. 健保記錄資料庫回應後將最新的健保記錄傳送給我們的 Server。
8. Server 得到這三種資料後，將這些記錄匯進到資料庫中，而 Matlab 會將其套入我們所 Training 的 model 中，並將預測結果回傳給使用者的瀏覽器程式，使用者就可以得到他想知道的最新訊息。

參考文獻

- [1] 王紹漢 “環境污染與健康” 科技圖書 1994 Aug.。
- [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 1994 Sep.，pp23-58。
- [3] 王姿乃 “氣喘之環境與遺傳因素研究” 高雄醫學大學醫學研究所博士論文 2000。
- [4] 王旭丹 “氣喘中西醫雙效自療手冊” 旺文股份有限公司 2000 Sep.，pp.25-38。
- [5] “中央氣象局”網站 <http://www.cwb.gov.tw/>。
- [6] 朱慧德、周卓定 “以遠端資料挖掘建構虛擬資料倉儲” 中華管理評論 2000 May，Vol. 3，No.2 pp.69-81。
- [7] “行政院環保署”網站 <http://www.epa.gov.tw/>。
- [8] 吳明進 “臺灣北部地區降雨氣候之研究” 國立臺灣大學大氣科學研究所碩士論文 1994。
- [9] 周政宏 “神經網路—理論與實務” 松崗電腦圖書 1996 Feb.，pp43-69。
- [10] 林煜棋 “鋪面道路車行揚塵特性與排放係數之建立” 國立中興大學環境工程學系碩士論文 1998。
- [11] 林信忠 “資料萃取法在健保費用查核之研究” 全國管理碩士論文集 1999。
- [12] 林能暉、彭啟明、陳進煌、陳靖沅 “臺灣酸雨之研究：源與受體關係” 第六屆全國大氣科學學術研討會論文集 1999。
- [13] 周昌宏、許晃雄、陳正達、柯文雄、鄒治華 “臺灣環境變遷與全球氣候變遷衝擊之評析—氣候 (Environmental Change of Taiwan and The Assessment of Vulnerability to the Impacts of Global Change-Climate)” 行政院國家科學委員會專題研究計劃成果報告。
- [14] 林文修 “Data Mining 探索” (上)(下) 決策論壇 第 13、14 期，<http://www.gss.com.tw/eis/13/mining.htm>,<http://gss.com.tw/eis/14/mining2.htm>
- [15] “空氣品質保護與噪音管制處” 網站 http://www.epa.gov.tw/airpollution/空保處_0205/default.htm
- [16] 張家誠、王立 “氣候變化四問” 明文書局 1994。
- [17] “國家衛生研究院”網站-全民健康保險學術研究資料庫<http://www.nhri.org.tw/nhird/>
- [18] 斯華齡 “電腦人腦化 神經網路—第六代計算機” 松崗電腦圖書 1993 Aug.，pp.79-92。
- [19] 喬安·克洛著，陳芳智譯 “過敏 305 問” 原水文化 2000 Dec.，pp45-91。
- [20] “燃燒的地球” 網站 <http://www.netvigator.com.tw/~a621027/first.htm>
- [21] A.Vellido，P.J.G Lisboa，J.Vaughan “Neural Networks in Business: a survey of applications

- (1992-1998)” Expert Systems with Application 17 (1999) pp.51-70。
- [22] Bo K. Wong, Thomas A. Bodnovich, Yakup Selvi “Neural Network applications in the literature (1988-95)” Decision Support Systems 19 (1997) pp.301-320。
- [23] Cabena , P. P. O. Hadjinian , R. Stadler , DR. J. Verhees , and A. Zanasi (1997), Discover- ing Data Mining from Concept to Implementation , Prentice Hall , pp.12 。
- [24] Michael J.A. Berry Gordon Linoff “Data Mining A Hands-on Approach for Business Professionals”. Data Mining 。
- [25] Nicolas Pasquier, Yves Bastide, Rafik Taouil and Lotfi Lakhhal “Efficient Mining of association Rules Using closed Itemset Lattices” Information Systems 1999 Vol. 24 No.1, pp.25-46 。
- [26] Rakesh Agrawal , Tomasi Imielincki , Aran Swami “Mining Association Rules Between Sets of Item in Large Database” ACM 1993 。
- [27] Wardlaw AJ. “The role of air pollution in asthma.” Clin Exp Allergy 1993 ;23:pp.81-96 。
- [28]張瑋倫，「資料挖掘與顧客關係管理之研究」，1999 中華民國科技管理論文集。
- [29]李紹倫、陳富中，「資料發掘在信用卡之應用」，1999 中華民國科技管理論文集。
- [30]Dr. Akeel Al-Attar (1998) ,”Data mining — Beyond Algorithms ” , <http://www.attar.com/tutor/mining.htm>
- [31]郭展盛，「數量化資料模糊挖掘模式之研究」，88 年全國管理碩士論文集。
- [32]藍中賢，「結合模糊集合理論與貝式分類法之資料探勘技術」，89 年中山大學資管學系研討會論文集。
- [33]邱顯貴，「以資料發掘經由知識管理建構顧客關係管理模式之研究」，89 年中山大學資管學系研討會論文集。
- [34]H. Michael Chung , Michael Manninon, ”Introduction to Data Mining and Knowledge Discovery”,1998.
- [35]陳文地，「資料探勘應用於犯罪分析模式建構之研究」，88 年世新大學資訊管理系研

討會論文 p95-102。

- [36]曾詠淑，「運用資料挖掘技術預測救護車服務量」，成功大學工程科學系碩士論文，87 年 6 月。
- [37]沈維明，「資料挖掘之關聯性法則架構以零售業目標行銷為例」，1999 年中華民國科技管理論文集。

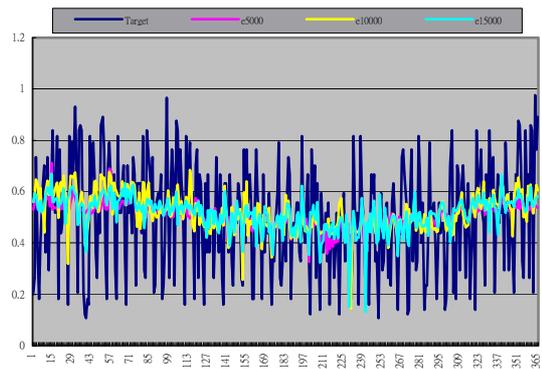


圖 1：北市空污+氣象台北、竹子湖

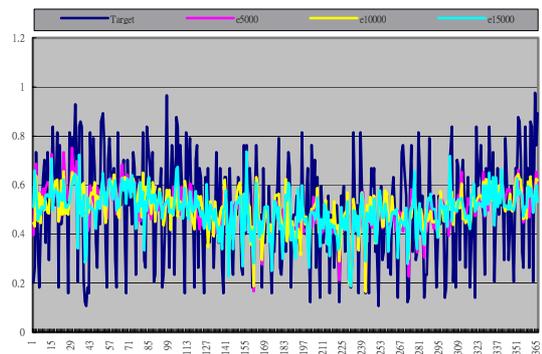


圖 2：北市空污+氣象竹子湖

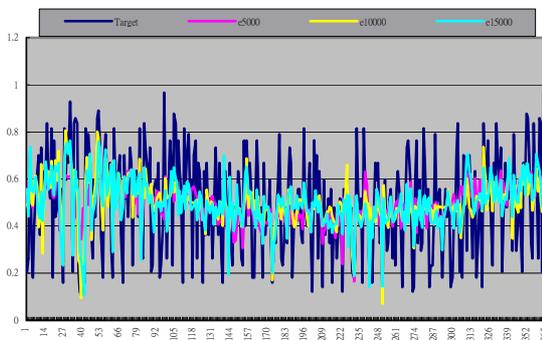


圖 3：北市空污+氣象台北

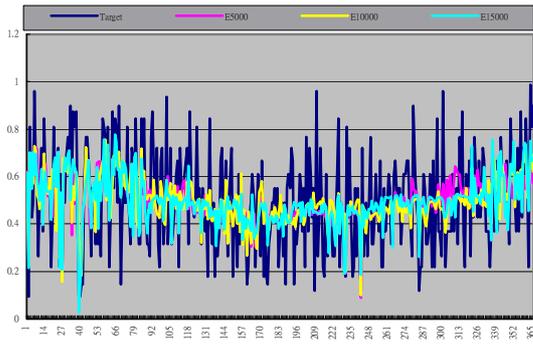


圖 4：北縣空污+氣象台北、淡水

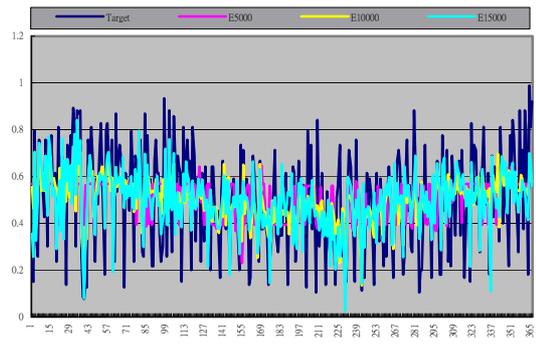


圖 8：訓練模型含次日記錄

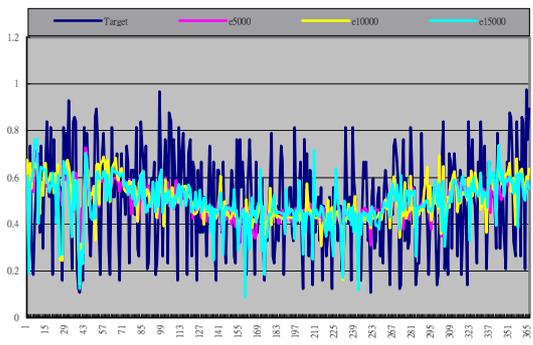


圖 5：北縣空污+氣象台北

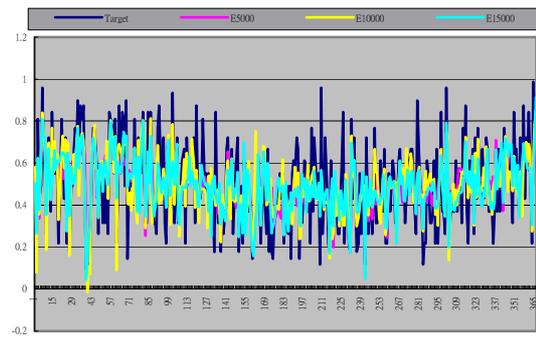


圖 9：訓練模型含後二日記錄

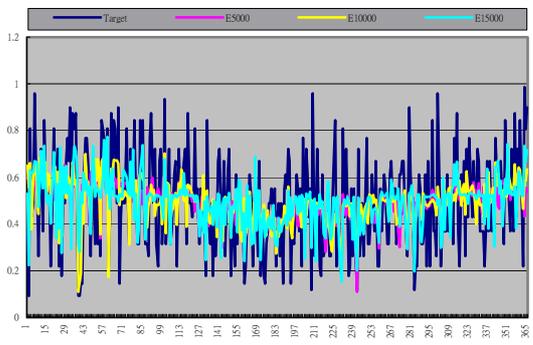


圖 6：北縣空污+氣象淡水

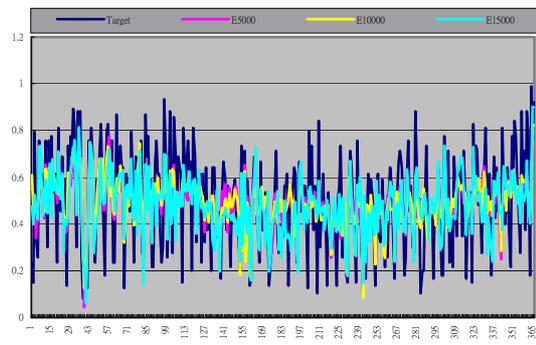


圖 10：訓練模型含後三日記錄

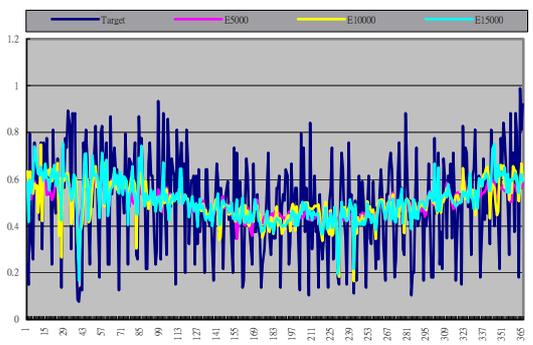


圖 7：大台北地區