第三章 研究方法

本研究是採用監督式知識發掘中的倒傳遞類神經網路(Back-propagation Network, BPN),作為主要的研究方法。BPN 在學習的方法上係屬於監督式學習法;而在架構上是多層前饋式架構。本章主要概述 BPN 類神經網路架構、運算過程及使用 BPN 網路時需注意的事項。

3-1. 倒傳遞類神經網路簡介

自感知機(Perceptron)由 F. Rosenblatt 在 1957 年提出後,由於缺乏隱藏層的學習演算法,導致學習能力大受限制。1969 年 M.L. Minsky 更提出感知機無法解決互斥或(exclusive OR, XOR)問題。

1974 年 P. Werbos 在其博士論文中首次提出利用隱藏層提高類神經網路的學習能力。1985 年,D. Parker 與 D. Rumelhart, G.E. Hinton 及 R.J. William 再次提出倒傳遞類神經網路的論文。從此,BPN 模式成為應用最廣泛、使用頻率最高的方法。

3-2. 倒傳遞類神經網路的架構

倒傳遞類神經網路為一多層前饋式結構的神經網路,網路架構如[圖 3-1],包含輸入層、輸出層及隱藏層。描述如下:

1. 輸入層：呈現類神經網路的輸入變數，視問題決定處理單元的數目。資料在輸入之前，必須先經過整理成有用的資訊後，才可成為輸入變數。使用線性轉換函數，即 f(X) = X。
2. 隱藏層：呈現各處理單元間交互影響的狀況。其處理單元數目與層數，並無固定的模式，需以試驗的方式來決定最佳的數目與層數。使用非線性轉換函數。
3. 輸出層：表現網路的輸出變數，其處理單元的數目亦視問題而決定。也是使用非線性轉換函數。

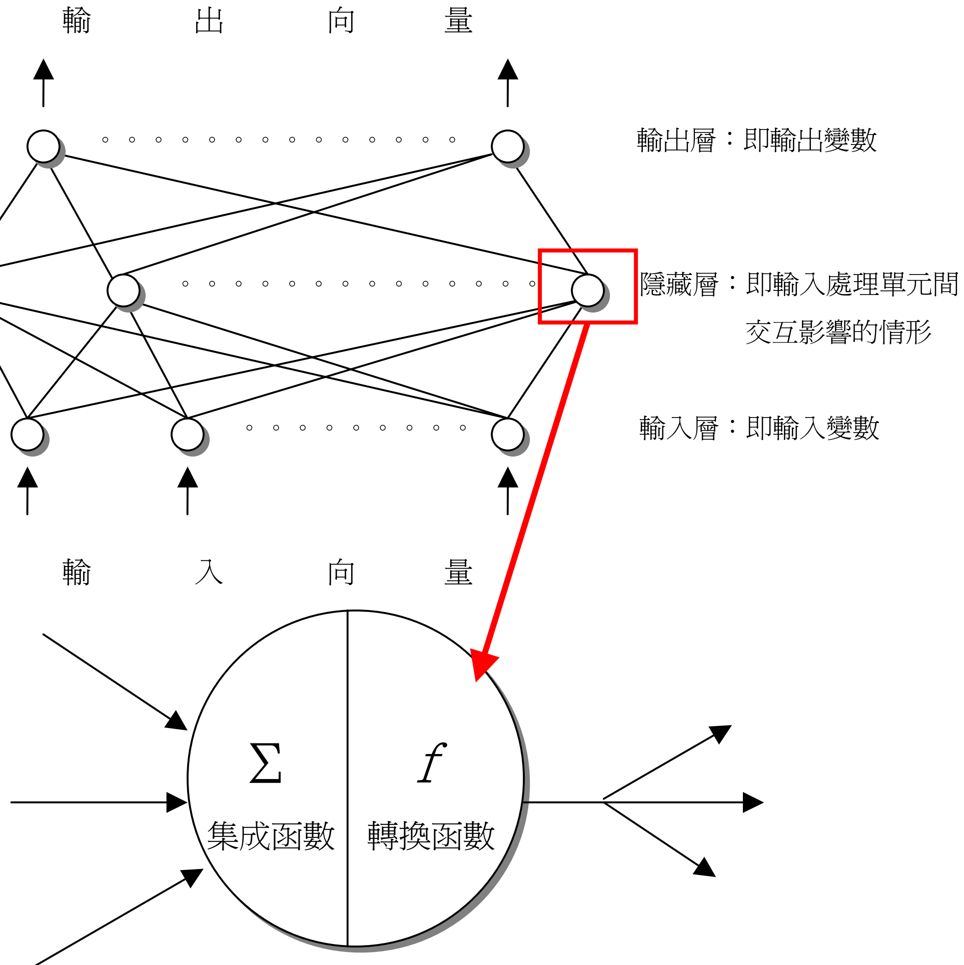

圖 3-1、倒傳遞神經網路架構圖
資料來源:本研究整理

倒傳遞網路常使用的非線性轉換函數為：雙彎曲函數(sigmoid function)[圖 3-2A]。

$$f(x) = \frac{1}{1+e^{-x}} \tag{3-1}$$

或雙曲線正切函數(Hyperbolic tangent function)[圖 3-2B]。

$$f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{3-2}$$

這種函數當自變數趨向正負無限大時，函數值將趨於常數，其函數值域在[0, 1]之間。

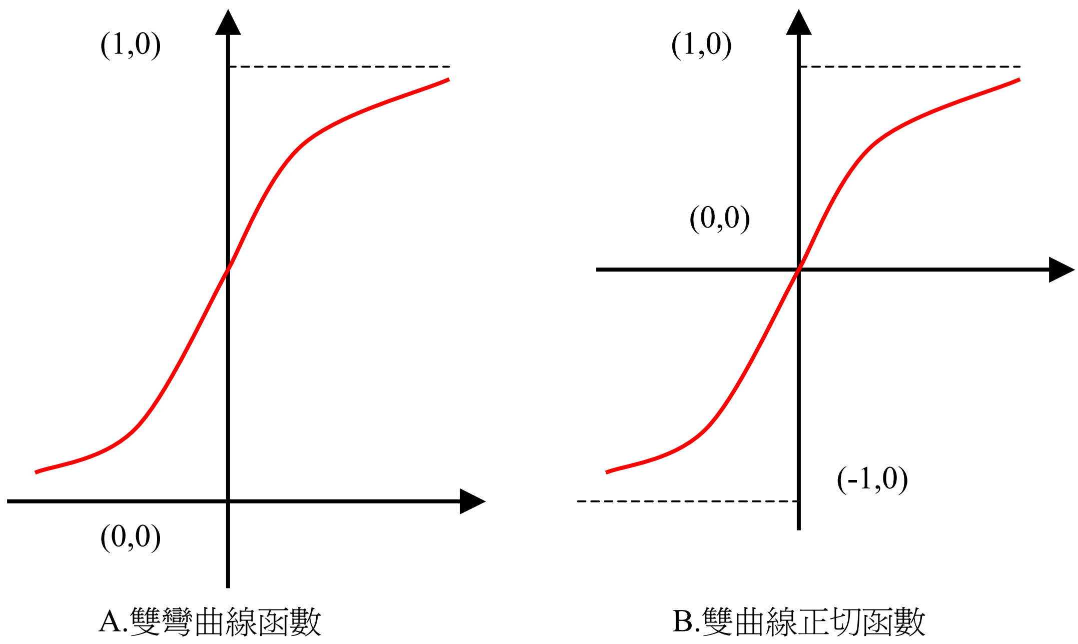

圖 3-2、非線性轉換函數
資料來源：葉怡成[1997]

3-3. 倒傳遞類神經網路的演算過程

茲以一個具有單隱藏層的層狀神經網路為範例，整理並說明倒傳遞類神經網路如何將一組訓練範例中的輸入值，與一組目標輸出值，藉由修正網路各處理單元的連結加權值和門限值，進而達到學習目的。

倒傳遞網路中，第 n 層的第 j 個單元的輸出值為第 n–1 層單元輸出值的非線性函數：

$$A_j^n = f(net_j^n) \tag{3-3}$$

其中

$$net_j^n = \text{集成函數} \quad = \sum_i W_{ij} A_i^{n-1} - \theta_j \tag{3-4}$$

根據(3-3)與(3-4)，假若應用訓練範例的輸入處理單元輸入值為{X}，則隱藏層隱藏處理單元的輸出值{M}如下:

$$M_k = f(net_k) = f(\sum W_{ik} X_i - \theta_k) \tag{3-5}$$

其中

(1) M_k 為隱藏層第 k 個隱藏單元的輸出值；
(2) f 為轉換函數
(3) W_{ik} 為第 i 個輸入單元與第 k 個隱藏單元間的連結強度；
(4) X_i 為第 i 個輸入單元的輸入值
(5) θ_k 為第 k 個隱藏單元的閥值

同理，應用隱藏層隱藏處理單元的輸出值{M}，推算出輸出層處理單元的推論輸出值{U}如下：

$$U_j = f(net_j) = f(\sum W_{kj} M_k - \theta_j) \tag{3-6}$$

其中

(1) U_j 為輸出層第 j 個輸出單元的推論輸出值；
(2) f　為轉換函數
(3) W_{kj} 為第 k 個隱藏單元與第 j 個輸出單元間的連結強度
(4) θ_j 為第 j 個輸出單元的閥值

根據(3-5)與(3-6)得到「推論的輸出值」與「目標的輸出值」相比較，可得到網路誤差。類神經網路即是利用此誤差來修正與調整連結中加權值的依據，以從訓練範例中建立系統模型。

監督式的學習網路旨在降低網路輸出單元目標輸出值與推論輸出值之差距。一般以能量函數來顯示學習的品質。

$$E = (1/2)\sum (T_j U_j)^2 \tag{3-7}$$

其中

(1) T_j 為輸出層第 j 個輸出單元的目標輸出值；
(2) U_j 為輸出層第 j 個輸出單元的推論輸出值；

透過網路的學習過程，經由不斷地修正網路連結上的加權值，將可使誤差函數達到最小值。為使誤差函數達到最小，一般是使用「最陡坡降法」。所謂「最陡坡降法」就是：每當輸入一個訓練範例，網路即小幅調整連結加權值的大小，調整的幅度和誤差函數對該加權值的敏感度成正比。就是誤差函數對加權值的偏微分值大小成正比。

$$\Delta W = -\eta \frac{\partial E}{\partial W} \tag{3-8}$$

其中 η 稱為學習速率(learning rate)，控制每次權值的修改的幅度。

針對隱藏層與輸出層間之連結加權值和輸入層與隱藏層間的連結加權值，可利用微積分中的連鎖率(chain rule)來導出。最後得到：

$$\Delta W_{ij} = \eta \cdot \delta_j^n \cdot A_i^{n-1} \tag{3-9}$$

此公式即為倒傳遞網路之關鍵。其中

(1) η 表示學習速率。

(2) δ_j 表示誤差量。

這種學習法則稱之為「通用差距法則」(General Delta Rule)。

倒傳遞神經網路，是透過一次又一次不斷反覆地以訓練範例學習，直至達到收斂，即可判斷所求之值是否接近預期的目標。為了檢驗學習的成果，在收集範例訓練資料時，必須先將資料分成兩個部分。一是正常的訓練用的資料，另一是測試使用的資料。並藉由以下兩種判斷基準，以檢測網路是否會達到收斂。

(1) 誤差均方根(Root of Mean Square, RMS)

針對函數型問題網路的誤差程度，可使用誤差均方根來檢測。

$$RMS = \sqrt{\frac{\sum_{p}^{M}\sum_{j}^{N}(T_{jp} - Y_{jp})^2}{M \cdot N}} \tag{3-10}$$

其中

T_{jp}= 第 p 個範例的第 j 個輸出單元之目標輸出值

Y_{jp}= 第 p 個範例的第 j 個輸出單元之推論輸出值

M = 範例個數

N = 輸出層處理單元數目

(2) 誤判率(Error Rate)

針對分類型問題網路的誤差程度，可使用誤判率來檢測。

$$\text{誤判率} = \frac{\text{範例總數} - \text{正判範例數}}{\text{範例總數}} \tag{3-11}$$

茲將倒傳遞網路演算過程整理如下圖[圖 3-3]：

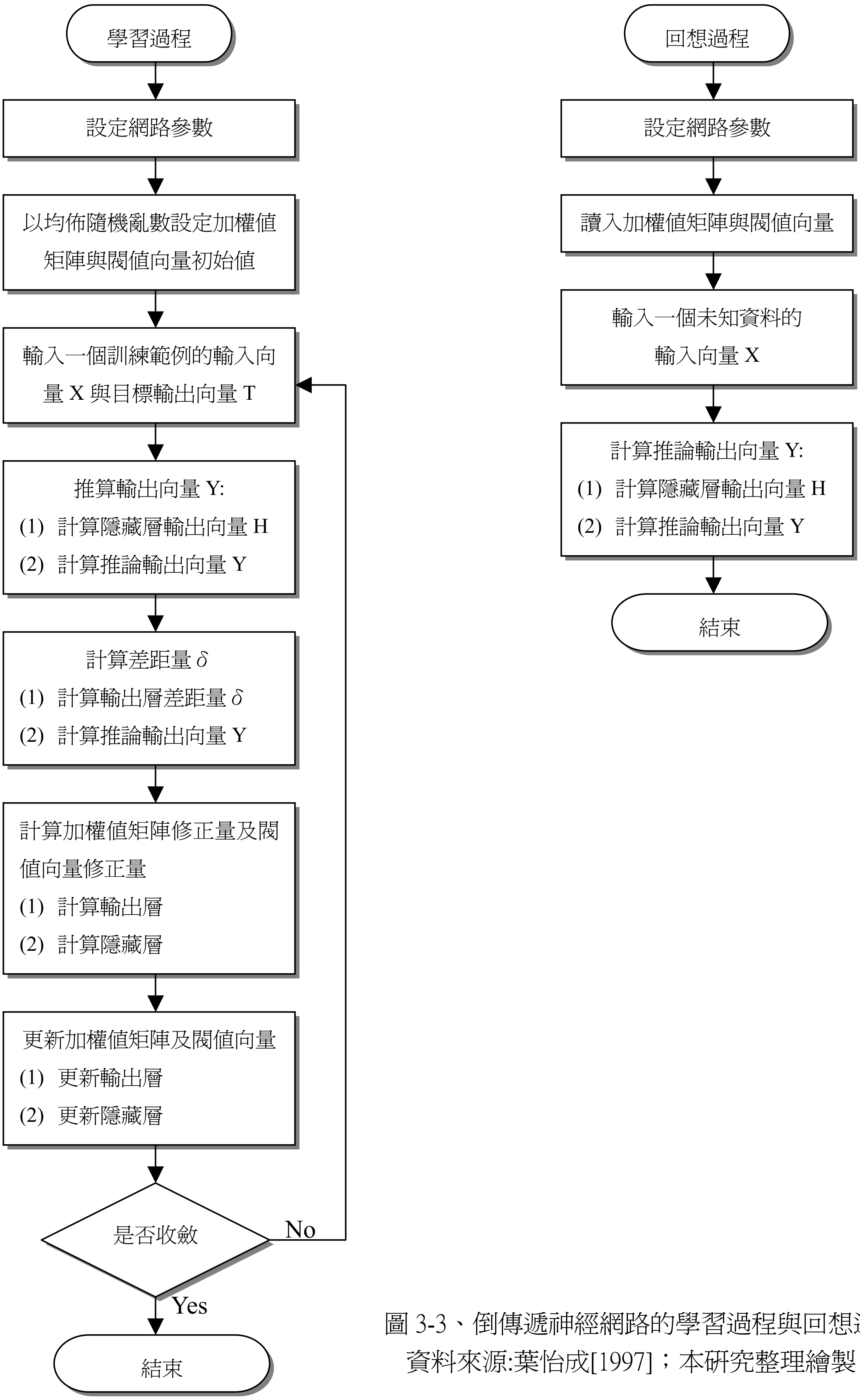

圖 3-3、倒傳遞神經網路的學習過程與回想過程
資料來源:葉怡成[1997]；本研究整理繪製。

3-4. 網路參數

在倒傳遞神經網路模式中，必須先設定相關的參數，例如：隱藏層層數、隱藏層處理單元數、學習速率及資料正規化。

一、隱藏層層數的考量

在倒傳遞神經網路中，隱藏層層數的設置，將會影響到網路執性的效能與結果分析的準確度。隱藏層若設置一~二層，有較佳的收斂的效果。若不設置隱藏層，將無法解決在輸入因子與輸出結果之間的非線性問題。另外一個問題是：有時在輸入訓練資料時，會發現資料彼此間呈現線性相關，這表示輸入因子與輸出目標之間，宜用線性迴歸分析，不適合使用類神經網路來求解。

反之，隱藏層數設置太多，會造成網路過度複雜，因而產生局部最佳化(即局部最小值)的問題。除此之外，也會造成隱藏層退化現象，使得收斂時間增長。

根據一般經驗，如果範例樣本數目少，雜訊的干擾較多且非線性程度較低的問題，使用一層的隱藏層即可。否則，就必須考慮是否要使用二層隱藏層。

二、隱藏層處理單元數目的考量

隱藏層處理單元數目的決定，可由以下兩種方式判斷是否符合「架構模式的需求」或「是否反應輸入向量與輸出向量之間複雜的映射關係」[郭益銘, 1999]。

1. 訓練誤差(training error)：即輸入訓練資料後，所得模擬結果之誤差，類似一般檢定模式。

2. 測試誤差(testing error)：即將未參與訓練之範例資料輸入後，所得模擬結果之誤差，類似一般驗證模式。

根據葉怡成[1997]選取隱藏層處理單元數目的選取原則：

1. 一般通則：

   (1) 平均法：簡單問題 =(輸入層處理單元數+輸出層處理單元數)/2

   (2) 總和法：一般問題 =(輸入層處理單元數+輸出層處理單元數)

   (3) 加倍法：困難問題 =(輸入層處理單元數+輸出層處理單元數)*2

2. 問題雜訊高，隱藏層處理單元宜少。

3. 問題複雜性高，非線性相關、交互作用程度高，隱藏層處理單元宜少。

4. 測試誤差遠高於訓練誤差，則發生「過度學習」現象，隱藏層處理單元宜減少；反之，則需增加。

三、學習速率的考量

在網路學習的過程中，常會出現以下兩種現象：

1. 減緩現象：此現象之發生，係因誤差函數的改變量漸趨減小，而造成速度減慢。此時，在網路的學習訓練上，可能會浪費相當多的時間，也可能落入局部最小值中。因此，學習速率應逐步遞增，才可收斂。

2. 跳出現象：由於學習速率過大，導致網路加權值修正過量，因而造成誤差函數值呈現上下起伏震盪，難以達到收斂之目的。應將學習速率放緩，才能完成收斂。

[表 3-1]是一般學習速率的建議設定值：

表 3-1、學習速率建議設定值一覽表

問題分類	初始值	折減係數	下限值
函數型	5.0	0.95	0.1
分類型	1.0	0.95	0.1

資料來源:葉怡成[1997]；本研究整理製表。

四、輸入資料正規化的考量

利用倒傳遞神經網路學習時，由於輸入因子的資料型態各有所異，因此應將資料先行正規化(normalize)至(0.1,0.9)的對應範圍間。其目的係在避免輸出訊號的飽和(saturation)現象。所謂飽和現象，就是因轉換函數微分值趨近於零，造成差距量 δ_j^n 亦趨近於零，使得處理單元將無法修正與其相連之連結加權值，因而無法完成收斂。

3-5. 倒傳遞神經網路應用於時間序列的預測分析

在日常生活中，有許多的預測問題是與時間序列有關。例如：股市指數、交易量及交易價格的預測，金融匯市的外幣匯兌預測、醫學微生物濃度培養預測、颱風行徑路線預測及農產品生長及產量收穫預測等。

這些與連續性時間有關的問題，由於缺乏數理的模式，很難從現有的理論或是模型中得到解答。為此，必須改用以往的歷史紀錄，或實驗的數據資料等，作為預測的基本資料。公式的推導及模型的建立，常會採用統計學中的「時間數列分析」［葉怡成, 1997］。

常見的時間數列分析方法，有以下數種：

1. 自動迴歸分析　　AR(p)
2. 移動平均分析　　MA(p)
3. 自迴歸-移動平均分析　　ARMA(p,q)
4. 自迴歸-求和-移動平均分析　　ARIMA(p,d,q)

以 p 階自迴歸分析 AR(p)為例，其基礎模型為：

$$X_t = a_0+a_1X_{t-1}+a_2X_{t-2}+\ldots+a_pX_{t-p} \qquad (3\text{-}12)$$

其中

X_t 是第 t 個時刻之變數值

X_{t-1} 是第 t-1 個時刻之變數值，於此類推。

a_0，a_1，a_2.... a_p 是自迴歸分析係數。

倒傳遞神經網路，亦可處理上述時間上的問題。轉換成類神經網路的運算方式，其中：[葉怡成,1997]

輸入變數：$X_{t-1}, X_{t-2}, \ldots, X_{t-p}$

輸出變數：X_t 之數據

若使用 BPN 網路來作時間序列的分析，應先由其中最早期的資料開始訓練。再依時間先後順序，依序將資料輸入至網路中運算。但使用於時間序列的網路模型，作為輸入因子的訓練資料並不都侷限在單一的時間序列中，大多都是連續性的時間數列[Berry and Linoff, 1997]。因此，以某一個特定的時間為基準點，將此基準點向前各若干天(視實際的需求而定)的資料，亦當成單獨的輸入因子，將可解決時間序列在倒傳遞神經網路上所面臨連續性資料輸入的問題。

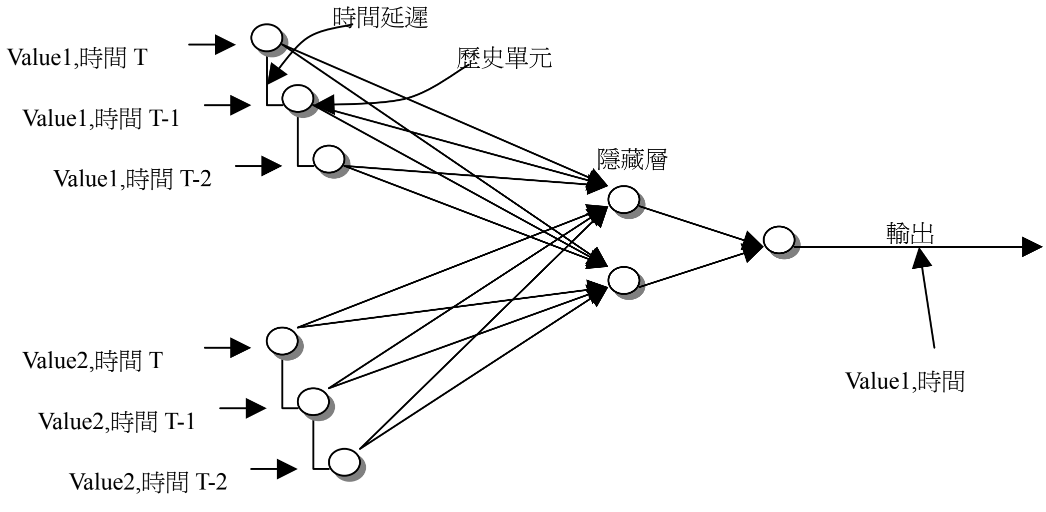

圖 3-4、倒傳遞神經網路應用在時間序列分析之結構圖
資料來源:Berry & Linoff [1997]