

改善多層次人工神經網路學習速度之探討

魏健宏* 洪玉城**

*成功大學交通管理科學研究所

**成功大學電機工程研究所

(收稿日期：89 年 1 月 13 日；第一次修正：89 年 4 月 24 日；
接受刊登日期：89 年 5 月 24 日)

摘要

近年來人工神經網路由於在辨識、分類、最佳化及推論等應用，均有不錯的效果，已引起人們廣泛的注意與研究。多層次人工神經網路之倒傳遞學習法則是一種監督式學習網路，在學習的過程中，需事先準備一群訓練樣本（亦即欲學習的知識庫），任意指定一組網路起始權值，以及若干參數，逐一輸入訓練樣本，計算輸出誤差量以調整網路權值，直到所有訓練樣本的網路輸出誤差，均小於某一容許值時，表學習完畢，可開始其應用。但在訓練過程中，傳統學習法則收斂速度極為緩慢，於是許多不同的改善方式陸續提出，本文將這些方法作一整理及扼要說明。本研究之前提係在不增加複雜度情形下，嘗試改善傳統倒傳遞學習法則，藉由找尋具有最大與最小誤差的輸出節點，並動態調整與該節點所連接權值的學習係數，以加快學習速度。本研究使用 Matlab 軟體撰寫程式模擬學習行為，並與傳統倒傳遞學習法則相比較，驗證在收斂速度有相當程度的改善。

關鍵詞彙：人工神經網路，多層次，倒傳遞學習法則，監督式學習

壹 前言

近年來，人工神經網路在管理科學領域的應用相當廣泛，在商業/財務、統計、作業研究/最佳化、交通運輸等方面已有不錯的效果。各項應用，說明如下：

- 商業/財務：公司/銀行破產預測、股票價格預測、債券評估、國家風險等級評估、資產分配、抵押保險、簽名確認、及商品交易等應用。
- 統計：分類、機率、迴歸、時間序列預測。
- 作業研究/最佳化：旅行銷售員問題 (Traveling Salesman Problem)、最短路徑、排序、線性規劃、組合最佳化、制定決策。
- 交通運輸：自動化高速公路匝道儀控系統、高速公路意外事件自動偵測模式、道路交通流量控制等應用。

如果以學習方式來區分，可將不同人工神經網路的數學模式 (Paradigms) 大約分為三種：

- 監督式學習網路 (Supervised learning network)
- 無監督式學習網路 (Unsupervised learning network)
- 聯想式學習網路 (Associative learning network)

監督式學習網路必須事先準備一組輸入與相對應明確的輸出樣本，作為其學習的訓練樣本。而無監督式學習網路，則無須準備一組明確的輸出，它直接從輸入樣本中擷取一些內在的規則或特徵，作為日後判斷其它樣本的依據。聯想式學習網路則類似於無監督式學習網路，但偏重於輸入不完整的資料時，可推論其原來完整的資料。三種不同的學習法則，各有不同的應用範例與不同的網路拓樸 (Topology)。而人工神經網路在統計領域應用上，約有七成使用倒傳遞學習法則 (Sharda, 1994)，促使本文針對倒傳遞網路 (Back propagation, 以下簡稱 BP) 各項特性來加以探討，並在不增加學習法則的複雜度下，提出了混合的學習方式，稱為 BP 最大誤差加速法，以求改善傳統 BP 學習速度。

BP 較完整的發展是由 Rumelhart et al. (1985) 所研究提出，是監督式學習網路的一種，允許多個隱藏層存在，其學習法則是以誤差能量函數 E (Energy function) 大小作為其收斂與否的指標。

假如吾人定義誤差能量函數為

$$E = \frac{1}{2} \sum_j (T_j - Y_j)^2$$

式中，

j ：輸出節點的數目

T_j ：真正所需要的輸出值

Y_j ：網路的輸出值

並以坡度下降法 (Gradient descent) 求其每次更新的權重值 ΔW ，則

$$W^n = W^{n-1} + \Delta W \quad (1)$$

$$\Delta W = -\eta \frac{\partial E}{\partial W} \quad (2)$$

式中，

η ：學習係數

$\frac{\partial E}{\partial W}$ ：誤差能量函數 E 對權值 W 偏微分，表示能量函數曲線的斜率大小

式(1)、(2)所代表的意義，可想像 ΔW 每次更新調整，均往較小的誤差能量函數 E 逼進，而學習係數 η 就表示每次更新 ΔW 步幅的大小。

由於網路各層不同的權值 W 對誤差能量函數 E 所構成的曲面為多維度複雜的曲面，因此學習規則採用坡度下降法時，誤差能量函數 E 易落於局部最小值不易跳出，致使較高的精確度無法達到。而 BP 的權值修正量 ΔW 是由輸出層依次往前修正，須隨著訓練樣本作遞迴的更新，直到收斂於某一精確度為止，因此學習速度緩慢。當外界的輸入值不適用人工神經網路時；或者人工神經網路的輸出需滿足某一特定輸出時，以及各項參數的設定等刻度化 (Scaling) 問題，均須在設計 BP 網路時作一整體性的考量。而且當 BP 人工神經網路學習完畢後，若有新的訓練樣本增加時，需整個重新學習一次，原來已學習的知識會有某種程度之變動。綜言之，傳統 BP 人工神經網路，具有下列主要的缺點：

- 可能會落於局部最小值
- 學習速度緩慢
- 刻度化問題
- 硬體電路上不容易實作
- 如要認知新的知識，須整個訓練樣本重新學習

為了解決上述缺點，於是出現了不同的研究方向與改善方法；有某些研究是針對局部最小值狀況作改善 (Baba, 1989)。有些則是在硬體實作上，針對晶片 (Chip) 面積、學習法則與精確性作研究 (Almeida & Franca, 1996; Dolenko et al., 1995)。也有些研究在討論如何保存既有的知識，而快速學習新的資訊 (Fu & Hsu, 1996)。但在改善學習速度上，就有相當多的改良式學習法則與不同的策略，在本文第二部份，即對這些改良式學習法則作一整體性、扼要的介紹。由於超大型積體電路 (VLSI) 設計持續的進步，使得大量神經元在同一個晶片的實作變為可行，不再侷限用軟體來模擬計算。因此，探討如何在

不增加學習法則的複雜度下，以方便硬體電路的實現 (implementation)，並且加以改善學習速度，為一重要的研究課題。在此研究目的下，本文第三部份提出了混合的學習方式，稱為 BP 最大誤差加速法，藉由簡易的判斷與運算，以求改善傳統 BP 學習速度。第四部份則用 Matlab 軟體寫了一個多輸入/多輸出、單隱藏層的 BP 最大誤差加速法，並與傳統 BP 學習法則作一比較，可驗證它具有較快的學習速度。最後，在第五部份則作了簡短的結論。

貳 各種加強BP學習效果的策略

改善 BP 學習速度的方法有相當多種，茲將較為人熟知的方法，簡述如下：

1. 慣性量 (Momentum term)：權值修正量修改為 $\Delta W^n = -\eta \frac{\partial E}{\partial W} + \alpha \Delta W^{n-1}$ ，多加了一個慣性項 $\alpha \Delta W^{n-1}$ ； α 稱為慣性係數。其物理意義可想為當這次的權值修正量 ΔW^n 與上一次的 ΔW^{n-1} 同向時，表示權重修正方向適當，但尚未跳離誤差能量函數 E 的低點，所以這次的 ΔW^n 可適度加大。反之，若 ΔW^n 與 ΔW^{n-1} 兩者方向相反，則表示已跳離 E 的低點，此次修正方向應調整且修正量應減小。但若 α 值太小，改善收斂速度效果有限； α 值如太大，則在收斂過程會有振盪產生，反而不利收斂速度。故選擇恰當的 α 值，需經由實際的訓練樣本測試才能決定其最佳值。
2. 批次學習 (Batch learning)：全部訓練樣本訓練一遍後，才更新權值一次。如此可避免個別訓練樣本偏誤，造成誤差能量函數 E 往錯誤的方向逼進，並可減少權值更新的次數，節省運算時間。
3. 採用不同的轉換函數：除了傳統的 sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ 外，尚有雙曲線函數 $\frac{e^x - e^{-x}}{e^x + e^{-x}}$ 、多項式曲線 e^{-x^2} 、 $\log(1 \pm x)$ 等，可應用於不同類型的訓練樣本。
4. 不同的誤差函數：如 $E = \frac{1}{n} \sum (T - Y)^n$ ，最小誤判 (Minimum misclassification error, MME) 由 Zahorian (1994) 提出。
5. 動態調整學習係數 η ：每一個 PE 有個別的學習係數 η ，檢查 δ^n 與 δ^{n-1} 是否同向，假如為真，導致學習係數 η 需增大；否則學習係數 η 需減小或不變。delta-bar-delta (DBD) (Jacobs, 1988) 與改良的 DBD 法 (EDBD) (Minia &

Williams, 1990) 均採用此法提昇訓練效果。

6. 動態調整轉換函數斜率：可加快學習速度與網路推理能力 (Generalization)，由 Tang & Kwan (1992) 提出。
7. 資料事先處理法：把訓練樣本先用數學方法處理、轉換，使其對 PE 所造成的激發 (Trigger) 較大，亦即把訓練資料空間換置在轉換函數斜率較大之處，以使得收斂速度加快 (Kwon & Cheng, 1996)。
8. 順向學習方式：把 BP 網路由後往前計算誤差、更新權值的方式，改為由前往後順向學習，可減少網路實作的複雜性 (Petridis & Paraschidis, 1995)。

其它尚有 net prune (Sietsma & Dow, 1987)、supplementary learning (Apolloni et al., 1990)、stochastic technique (Barnard, 1992)、fast BP (Samad, 1988)、cascade-correlation learning (Hwang et al., 1996)、layer-by-layer optimizing (Wang & Chen, 1996) 等方法。

上述的每一種方法，都有其特殊的演算方式與目的，有些是以加快其收斂速度為主，如項次 1、3、5、6、7、supplementary learning、fast BP；有些是以加強其精確性為主，如項次 2、4、net prune；有些是以尋找整體的最佳化為出發點，如 stochastic technique 及 layer-by-layer optimizing；而 cascade-correlation learning 為動態調整隱藏層數目的學習方式。下節即以動態調整學習係數 η 作為基礎，配合最大能量誤差，建立另一型演繹法—BP 最大誤差加速法，以求改善 BP 的學習速度，其演繹法詳見下一部份。

參 BP最大誤差加速法

本研究基本的概念是找出一簡單實用的判斷法則，有效地調整學習係數 η ，以求加速學習速度。當一 BP 網路共有 p 個輸出時，假使某一輸出節點 m ，在學習過程中有最大的學習誤差 E_m ，為求加快學習速度，該節點所連接的權值修正量 ΔW 應較大；相反的，若某一輸出節點學習誤差已經小於可容忍誤差時，該節點所連接的權值修正量 ΔW 應微調即可，以免跳出最佳值域。但由於轉換函數 sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ 在 x 的絕對值較大時，轉換函數 $f(x)$ 趨近於飽合，其微分值趨近於零，使得權值修正量 ΔW 變得較小，影響其收斂速度。故當學習係數 η 調大時，應在轉換函數微分後加一微小項成為 $[f'(x) +$

unsat] , unsat 為一固定小數值，以防止 $\Delta W \approx 0$ 。茲將 BP 最大誤差加速法流程整理如下。

一、符號與名詞定義

η_m^n : 第 n 次訓練時，各權值 W 連到輸出節點 m 的學習係數。

E_m^n : 輸出節點 m ，第 n 次訓練的學習誤差值。

unsat : 為一固定小數值，以下模擬時，其值為 0.05。

$E_{tolerance}$: 預定的容忍誤差，設定較低的容忍誤差表示精確度要求高，反之亦然。

收斂 : 定義為輸出層的誤差能量函數 E 小於預定的容忍誤差 $E_{tolerance}$ ($E < E_{tolerance}$)。

二、學習法則

Step 1 : 給一初始學習係數 η_{org}

Step 2 : 找出最大與最小誤差值的輸出節點 $E_m = \max\{E_i, i=1 \sim p\}$; $E_t = \min\{E_i, i=1 \sim p\}$

Step 3 : 假使 $(E_m^n - E_m^{n-1}) \leq 0$, 則學習係數 η_m^{n+1} 調大 ($\eta_m^{n+1} = 2\eta_{org}$) 及轉換函數微分後加一微小項 [$f'(x) + \text{unsat}$] ; 否則學習係數 η_m^{n+1} 變小 ($\eta_m^{n+1} = 0.5\eta_{org}$) , 及 $\eta_{\neq m}^{n+1}$ 不變。

Step 4 : 假使 $E_t^n \leq E_{tolerance}$, 則將學習係數 η_t^{n+1} 變小 ($\eta_t^{n+1} = 0.5\eta_{org}$) , 而 $\eta_{\neq t}^{n+1}$ 不變 ; 否則學習係數 η_t^{n+1} 不變。

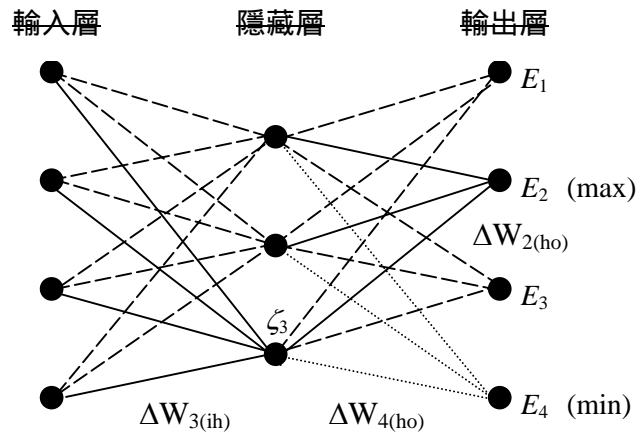
Step 5 : 重覆上述學習法則，直至收斂或最大的學習次數。

由於倒傳遞人工神經網路為多層的拓樸架構 (Multi-layer topology) , 從輸出層學習誤差 E 可反向依序推論出每一隱藏層的誤差。本法則專注改良輸出端學習誤差 E 的變化趨勢與相互大小關係，選擇具有最大學習誤差 E_m 的輸出端作增強處理，而不再依序求取隱藏層中具有最大誤差之節點 (如 DBD 方法) 作相同的權值增強處理，其它層的更新模式則維持不變。基本上，本法則的學習過程是傳統 BP 的延伸，其理論依據並未砥觸原數學推論，僅在具有最大學

習誤差 E_m 與最小學習誤差 E_t 的輸出節點上，作學習係數的加強與減弱，而並未修改能量誤差函數定義或修改在隱藏層權值的更新法則。其精神在於，如果預知權值的修改方向是正確的（能量函數 E 值是持續地往低誤差方向逼近），且此輸出節點是具有較大的輸出誤差，則連接此輸出節點的權值應較其它節點的權值要有更大幅度的更新（見 Step 3）。或者，某輸出節點的輸出誤差已經小於所容許的誤差時，表示所連接的權重值已經穩定於訓練樣本所呈現的特性，此時不應再大幅度更改（見 Step 4）。而權重更新的數學計算模式，仍然是依據傳統 BP 的坡度下降法。但其加強或減緩更新的判斷法則，則較 DBD 方式簡便。而且只有在輸出層作觀察，此過程係基於下列兩個理由：

1. 倒傳遞人工神經網路最外層輸出，原本是分析者欲直接觀察的輸出端點，在此取得學習誤差 E ，不需經繁複的計算，在架構上較具彈性與簡便。
2. 在局部最佳狀態下，不希望影響其它已趨於穩定的權值，如圖一說明。例如圖一輸出層最大學習誤差是 E_2 ，最小學習誤差是 E_4 ，此學習法則是要增強所有連接 E_2 的權值變化 $\Delta W_{2(ho)}$ ，而欲保留連接 E_4 的權值 $W_{4(ho)}$ 。假使繼續推論到隱藏層中，具有最大誤差 ξ_3 而增強 $\Delta W_{3(in)}$ 時，會間接影響到原先欲保留連接 E_4 的權值 $W_{4(ho)}$ ，而破壞局部最佳狀態，甚至影響整體收斂效果。

為求簡便展示初步構想，上述 Step 3 與 Step 4 學習係數 η_m^{n+1} 變大與 η_t^{n+1} 變小，各調整為原來 2 倍與 1/2 倍。亦即在整個學習過程中，學習係數不外乎 η_{org} 、 $2\eta_{org}$ 與 $0.5\eta_{org}$ 三種，所以可在下一節結果得知，設定不同的學習係數 η_{org} 對學習效果有顯著的影響。由於學習係數只有事前設定的三種，可事先儲存備用，不像 DBD 法則每次調整學習係數時，調整公式可能相當複雜，均需重新計算。故本方法在學習過程，計算較為簡潔省時。增強與減弱因子各設為 2 倍與 1/2 倍，是為了有效加快學習速度，並可藉由調整初始學習係數 η_{org} 的大小，來適應不同應用範例的需求。此外，2 倍與 1/2 倍在硬體實作（VLSI design）上較為簡便，電路亦可較容易地設計，數位硬體架構上可將原值左移或右移一個位元（bit），即完成必要之改變。如此，不論在硬體實作上與軟體的模擬，各有其便利性。



圖一 單一隱藏層 BP 網路架構

肆 模擬結果與探討

本次研究的模擬環境是以個人電腦為作業平台，並且利用 Matlab 4.0 寫了一個多輸入多輸出單隱藏層，各權值 W 具有隨機起始值的程式。

此處以花種分類的資料 (葉怡成, 1995)，作為測試「BP 最大誤差加速法」與「傳統 BP 學習法則」性能好壞的訓練樣本 (75 個 4-input/3-output 的樣本)。此訓練樣本是測量不同花種間，其萼片長、萼片寬、花瓣長、花瓣寬之間的關係。由於人工神經網路各權值 W 及誤差函數所構成的是多維度複雜曲面，不同的初始權值分佈與隱藏層節點多寡對於學習速度有密切的關係，為了能客觀評估在不同情形下的學習速度及節省模擬時間，本研究在不同的初始參數組合下隨機產生 50 組起始權值，分別對兩種學習法則進行模擬。各項參數為：

- 網路架構取 4-2-3 與 4-3-3 兩種
- 初始學習係數 η 取 0.1、0.5 或 0.8
- 容忍的誤差值 $E_{tolerance}$ 取 0.1、0.05 或 0.01
- 最大學習次數在 $E_{tolerance}=0.1、0.05$ ， $\eta=0.1、0.5$ 時為 300 次
 $E_{tolerance}=0.1、0.05$ ， $\eta=0.8$ 時為 150 次
 $E_{tolerance}=0.01$ ， $\eta=0.1、0.5、0.8$ 時為 1000 次

表一 「傳統 BP 法則」與「最大誤差加速法則」各項模擬結果

參數		4-2-3 架構		4-3-3 架構	
E_{tol}	η_{org}	傳統 BP 法則	最大誤差加速	傳統 BP 法則	最大誤差加速
0.1	0.1	292/ 0/35	219/44/ 6	284/ 0/24	206/48/ 2
	0.5	151/20/ 0	163/30/ 9	129/12/ 0	132/36/ 4
	0.8	137/ 9/31	126/20/30	133/12/26	119/25/21
0.05	0.1	X/ 0/50	X/ 0/50	X/ 0/50	X/ 0/50
	0.5	231/18/10	222/29/15	196/13/ 0	193/36/ 7
	0.8	X/ 0/50	143/14/36	149/ 2/48	146/15/35
0.01	0.1	X/ 0/50	X/ 0/50	X/ 0/50	X/ 0/50
	0.5	X/ 0/50	555/34/16	903/ 1/45	530/43/ 7
	0.8	751/23/ 4	408/25/25	704/13/ 5	401/36/13

註：表中數值 a/b/c：

- a，表示收斂時訓練次數之平均值，用於評估學習速度，其值愈小學習速度愈快；X 表在規定的學習次數內仍未收斂。
- b，表示快於他法的模擬次數，用於評估在不同的初始權重值下，學習的穩定度。
- c，表示發散或到達最大學習次數仍未收斂的模擬次數，此項數值並未併入 a 項求取平均值。

各項模擬結果如表一所示，其中第一列 219/44/6 表示「最大誤差加速法」在學習係數 $\eta=0.1$ 、 $E_{tolerance}=0.1$ 、50 組隨機起始權重值模擬下，平均而言學習 219 次即達到收斂標準，其中有 44 組快於傳統 BP 法則，但有 6 組在最大學習次數限制內（在此為 300 次），未能收斂。從表一可觀察，「最大誤差加速法」在精確度要求較低時（ $E_{tolerance}=0.1$ 、 $\eta=0.1$ ）收斂次數改善效果較佳，而且收斂平均值（a 值）較不受架構變化的影響，學習速度較穩定。

根據以上的模擬結果，以學習係數 η 及容忍誤差 $E_{tolerance}$ 的大小作為評比要素，比較「最大誤差加速法」與「傳統 BP 法則」兩者學習速度（收斂次數），如表二所示。在學習係數 η 小，容忍誤差 $E_{tolerance}$ 大（精確度要求低）的情況下，「最大誤差加速法」有明顯的改善效果，平均約節省 27% 的學習循環次數。在學習係數 η 大，容忍誤差 $E_{tolerance}$ 小（精確度要求高）的情況下，平均約快 43% 的學習循環次數，但在規定的學習次數內，亦有較大的不收斂次數，較為不穩定。在學習係數 η 小，容忍誤差 $E_{tolerance}$ 小的情況下，則無改善效果。其餘區間，學習速度有少許改善，但效果不明顯。

表二 「最大誤差加速法則」相對於「傳統 BP 法則」特性比較

容忍誤差\學習係數	小 ($\eta=0.1$)	中 ($\eta=0.5$)	大 ($\eta=0.8$)
大 ($E_{\text{tolerance}}=0.1$)	改善 27%，穩定	少許改善	少許改善
中 ($E_{\text{tolerance}}=0.05$)	績效相似	少許改善	少許改善
小 ($E_{\text{tolerance}}=0.01$)	績效相似	改善 41%，穩定	改善 43%，不穩定

從表一與表二可得知，「最大誤差加速法」在多數情形下，平均收斂次數與穩定度均比傳統 BP 法則為佳。在精確度要求高 ($E_{\text{tolerance}}=0.01$) 與適當設定學習係數 η 時，學習速度有明顯的改善。但在學習係數 η 設定太小而精確度要求高的情形下，改善效果不佳的原因，可能在學習過程中，輸出誤差 E_m 或 E_r 常出現在少數輸出節點，如此第四節內 Step 3、Step 4 步驟僅在少數輸出節點發生作用，使得整體的輸出誤差 E 容易陷入局部凹點，無法快速脫離而影響到學習速度，這或許是此學習法則的未來改善方向。此方法雖能改善學習速度，但針對不同的參數設定時，學習績效仍不夠穩定。而且，若是多個的隱藏層時，整體效果是否能有顯著改善，仍需再作模擬。但需模擬幾層的隱藏層、幾層才算具有代表性、何種訓練樣本才足夠代表多樣性，這是我們模擬的困難與研究需加強的地方。但此研究，能在不增加學習規則的複雜度，並方便使用硬體電路實現的要求下，藉由簡易的動態調整，完成加快學習速度的目標，亦已達成預估的構想。

一般採取動態調整學習係數，以增快學習速度的學習策略較著名的有 DBD 方式，但由於其輸出處理單元及隱藏處理單元均需一個獨立的學習係數，而且可調整的參數眾多，雖可滿足不同類型訓練樣本的應用，但相對的，也招致調諧各項參數的困擾。若有足夠運算資源以嘗試錯誤方式，選擇 DBD 方法所需的一組最佳參數，其收斂速度確優於本文所提出方法。但由於其各項參數調整與判斷的依據 (criteria) 較為繁雜，不容易將其精神用硬體電路來實現，而本方法嘗試僅改變輸出層的學習係數，來加快整體的學習速度，卻不會將學習法則複雜化，表三為此兩種動態學習法則的比較。

此次研究是以簡單地調整學習係數 η 為重心，如能配合不同的轉換函數或調整其斜率，將更具改善的效果。但話說回來，改善學習速度的目的，不外乎是要節省學習的時間，如果把一種學習法則所需執行 n 次遞迴 (iteration)，每次遞迴執行 t_1 時間，共消耗 $n*t_1$ 時間的演算法，改成另一種較複雜的演算法，收斂次數節省至 m 次 ($m < n$)，但每次執行 t_2 時間 ($t_2 > t_1$)，則共執行 $m*t_2$

時間，假使兩者演算法的總共執行時間相差不多 ($m*t_2 \approx n*t_1$)，則是否有其需要改成較複雜的演算法，就需全面性的評估了，包括考量硬體實作上困難度、精確性的高低及是否適合不同類型的訓練樣本等因素。

表三 「最大誤差加速法則」相對於「DBD 法則」 特性比較

比較項目 \ 方式	最大誤差加速法則	DBD 法則
1.主要的目的	不增加學習法則的複雜度，改善學習速度	全面改善學習速度
2.需調整的節點	在輸出層具有最大誤差與最小誤差的節點	在輸出層及隱藏層的各個節點
3.需規劃與調整的參數	學習係數	學習係數、凸出係數、增常數、減常數
4.學習係數更新方式	藉由設定 η_{org} ，動態調整 η_{org} 、 $2\eta_{org}$ 與 $0.5\eta_{org}$	每次均需重新計算後，動態更新連續值 η
5.學習法則運算複雜度	較簡易	較複雜
6.硬體電路實現	較容易	較複雜
7.學習速度加快的效果 (與傳統 BP 比較)	快、穩定度不足	最快、穩定度高 (需各項參數設定在最佳組合情形下)

伍 結論

本文章扼要介紹人工神經網路種類及其應用，並討論倒傳遞網路在應用上所遭遇的問題，並參酌文獻中各種不同加強 BP 學習效果的策略，提出「BP 最大誤差加速法」，藉著找尋最大誤差能量的輸出節點，簡單的改變學習係數 η (為原來的 2 倍或減低 0.5 倍)，來增強局部的權重值變化。並且利用 Matlab 4.0 寫了一個多輸入多輸出單隱藏層的模擬程式，以花種分類的資料，測試「最大誤差加速法」在不同初始權重值下之效果。模擬的結果顯示以本研究所提出的學習法則，在不將原有的學習法則複雜化下，配以適當的學習係數 η ，可有效地加快 BP 的學習速度。

參考文獻

葉怡成，「類神經網路模式應用與實作」，台北：儒林書局，1995 年。

- Almeida, A. P. and Franca, J. E., "Digitally Programmable Analog Building Blocks for the Implementation of Artificial Neural Networks", *IEEE Transactions on Neural Networks*, Vol.7, No.2, 1996, pp.506-514.
- Apolloni, B. et al., "Diagnosis of Epilepsy via Backpropagation", *Proceedings of the International Joint Conference on Neural Network*, II, 1990, pp.571-574.
- Baba, N., "A New Approach for Finding the Global Minimum of Error Function of Neural Networks", *Neural Networks*, Vol.2, 1989, pp.367-373.
- Barnard, E., "Optimization for Training Neural Nets", *IEEE Transactions on Neural Networks*, Vol.3, No.2, 1992, pp.232-240.
- Dolenko, B. K. et al., "Tolerance to Analog Hardware of On-Chip Learning in Backpropagation Networks", *IEEE Transactions on Neural Networks*, Vol.6, No.5, 1995, pp.1045-1052.
- Fu, L. M. and Hsu, H. H., "Incremental Backpropagation Learning Networks", *IEEE Transactions on Neural Networks*, Vol.7, No 3, 1996, pp.757-761.
- Hwang, J. N. et al., "The Cascade-Correlation Learning: A Projection Pursuit Learning Perspective", *IEEE Transactions on Neural Networks*, Vol.7, No.2, 1996, pp.278-288.
- Jacobs, R. A., "Increased Rates of Convergence through Learning Rate Adaptation", *Neural Networks*, 1, 1988, pp.295-307.
- Kwon, T. M. and Cheng, H., "Contract Enhancement for Backpropagation", *IEEE Transactions on Neural Networks*, Vol.7, No.2, 1996, pp.515-524.
- Minia, A. A. and Williams, K. D., "Acceleration of Back-propagation through Learning Rate and Momentum Adaptation", *Proceedings of the International Joint Conference on Neural Network*, I, 1990, pp.676-679.
- Petridis, V. and Paraschidis, K., "On the Properties of the Feedforward Method: A Simple Training Law for On-Chip Learning", *IEEE Transactions on Neural Networks*, Vol.6, No.6, 1995, pp.1536-1541.
- Rumelhart, D. E. et al., "Learning Internal Representation by Error Propagation", in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, D. E. Rumelhart and J. L. McClelland, eds., Cambridge, MA, MIT press, 1986.
- Samad, T., "Back-propagation is Significantly Faster if the Expected Value of the Source Unit is Used for Update", *International Neural Society Conference Abstracts*, 1988.
- Sharda, R., "Neural Networks for the MS/OR Analyst: An Application Bibliography", *Interfaces*, Vol. 24, No. 2, pp.116-130, 1994.
- Sietsma, J. and Dow, R. J. F., "Neural Net Pruning- Why and How", *Proceedings of the International Joint Conference on Neural Network*, I, pp. 325-333, 1987.
- Tang, C. Z. and Kwan, H. K., "Convergence and Generalization Properties of Multilayer Feedforward Neural Networks", *IEEE International Symposium on Circuit and Systems*, pp. 65-68, 1992.
- Wang, G. J. and Chen, C. C., "A Fast Multilayer Neural Network Training Algorithm Based on the Layer-by-Layer Optimizing Procedures", *IEEE Transactions on Neural Networks*, Vol.7, No. 3, pp. 768-775, 1996.

Zahorian, S. A., "Comparison of Minimum Misclassification Error (MME) Networks with Least Mean Square Error (LMS) Networks", *Proceedings of the International Joint Conference on Neural Network*, 1994, pp.191-197.

Improving Learning Speed on Multilayer Artificial Neural Networks

CHIEN-HUNG WEI *, YU-CHERNG HUNG **

* *Department of Transportation and Communication Management Science, National Cheng Kung University*

** *Department of Electrical Engineering, National Cheng Kung University*

ABSTRACT

The performance of artificial neural networks for image recognition, sorting, optimization and generality during recent years is really impressed. More and more attention and research have been devoted to related topics. Multilayer neural networks with the back-propagation learning method become the most popular supervised learning networks. To initiate the learning process, a set of training samples (i.e., the knowledge to be learned) and arbitrary initial weights should be ready. An iterative procedure is conducted to compute network output discrepancy and to modify weights accordingly. These steps continue until the error is relatively small. However, the convergent speed of this conventional method is relative slow. There are some strategies successful in improving its performance. This article proposes a new approach. The underlying idea is to find the output nodes with the maximum and minimum error for adaptively modifying learning coefficients. Behavior of the modified learning rule is simulated and the results are compared with the conventional backward propagation learning rule. Some improving effects are observed. Finally, other approaches to further enhancing the learning speed of multilayer neural networks are discussed.

Keywords: artificial neural networks, multilayer, back-propagation learning, supervised learning

