# 1C－2 荷重空間内での誤差逆伝播学習法の振舞い

相澤雅彦* 山崎智子* 大照 完* 橋本周司**
*早稲田大学理工学部 **東邦大学理学部

## 1 はじめに

一般に誤差逆伝播学習によって、結合の重みは二乗誤差を極小にするようなものに収束することが指摘されている[1]。このとき訓練パターンをランダムに提示すると、学習過程での荷重ベクトルは荷重空間でランダムウォークすることになる。ここでは荷重空間内での荷重ベクトルの動きを確率過程論的に検討し、具体的な計算実験の結果と共に述べる。

## 2 荷重空間内でのポテンシャル関数

次の二乗平均誤差を最小化することが学習の目的である。

$$E=\langle E_k \rangle=\sum_k P(k)(y_k-d_k)^2 \qquad (2\text{-}1)$$

ここで$y_k$,$d_k$は入力パターンkに対する希望出力とネットワーク出力である。またP(k)はパターンkの発生確率である。

上記Eは荷重の関数であり、バックプロパゲーションは、このEの各荷重についての偏微分係数を0にするよう荷重ベクトル$\vec{w}$を変化させ、目的を達成しようとするものであるが、実際は$\langle E_k \rangle$を直接求めることはできないので、各パターンに対する二乗平均誤差$E_k$に付いて

$$\Delta w_m = -\varepsilon \frac{\partial E_k}{\partial w_m} \qquad m=1,2,\cdots,N \qquad (2\text{-}2)$$

を修正量として用いることになる。
従って、学習過程はポテンシャル関数$E_k(w_1,w_2,\cdots\cdots w_n)$がランダムに変化する荷重空間内での確率過程となる。

## 3 荷重ベクトルの確率過程

荷重ベクトル$(w_1,w_2 \cdots\cdots ,w_n)$をランダムなパターン提示によって修正するとき、各成分の増分の期待値$A_m$と分散$B_m$を求めると、

$$\begin{aligned} A_m &= \langle \Delta w_m \rangle = -\varepsilon \sum_k P(k) \frac{\partial E_k}{\partial w_m} \\ B_m &= \langle \Delta w_m^2 \rangle - \langle \Delta w_m \rangle^2 \\ &= \varepsilon^2 \sum_k P(k) \left( \frac{\partial E_k}{\partial w_m} \right)^2 - A_m^2 \end{aligned} \qquad (3\text{-}1)$$

$\varepsilon$が十分小さいとすれば、時刻tに荷重ベクトルが$\vec{w}$である確率密度U(t,w)は次のFokker-Planckの式を満足する。

$$\frac{\partial U}{\partial t} = -\sum_m \frac{\partial A_m U}{\partial w_m} + \frac{1}{2} \sum_m \frac{\partial^2 B_m U}{\partial w_m^2} \qquad (3\text{-}2)$$

上式より、パターン提示方法による収束速度の相違や平均学習時間等が推定できる。

## 4 具体例

### 4．1 ネットワークの設定

実際は3層対称構造のネットワークを用い、排他的論理和の学習について検討した。図1のように入力層と中間層の間、中間層と出力層の間の重みをそれぞれ一つずつ変数とした。

### 4．2 Rumelhartの学習[2]

学習過程を検討するためにRumelhartらの誤差逆伝播アルゴリズムを使用した。その更新式を（4－1）に示す。ただし、出力と希望出力の二乗誤差をポテンシャル関数E、重みをw、修正量を$\Delta w$、修正回数をkとする。

$$\Delta w(k+1)=-\varepsilon \frac{\partial E}{\partial w} + \alpha \Delta w(k) \quad \cdots (4\text{-}1)$$

ここで$\varepsilon=0.1$, $\alpha=1.0$で学習させたとき、各入力パターンごとのポテンシャル関数のグラフを図2、図3、図4、図5に示す。図中のグラフは二乗誤差の等しい点を結んだものである。また、種々の初期値から、パターンをランダムに1万回提示した時の収束過程を図6に示す。

これからわかることは、誤差が急激に減るところは幅が狭く、従ってここでは急速に学習が進むことを示す。。このことはsigmoid関数の形に起因するものと思われる。また、重みが大きくなるにつれてポテンシャル関数の傾きが小さくなる傾向があり、この場合、平坦なところが続いてしまい、重みが更新されにくくなることがわかる。

### 4．3 提示方法による修正量の分散

つぎに学習パターンの提示方法による分散の相違について考えてみる。

学習パターンの提示がランダムに行われると重みの変化の時間発展は確率過程となる。四つのパターンの提示確率が等しい場合、1ステップ当りの重みの修正量の平均Aおよび分散Bは次式で与えられる。

$$\begin{aligned} A_i &= -\varepsilon \frac{\partial \langle E \rangle}{\partial w_i} \qquad i=1,2 \\ B_i &= \varepsilon^2 \left[ \left\langle \left( \frac{\partial E}{\partial w_i} \right)^2 \right\rangle - A_i^2 \right] \end{aligned} \qquad (4\text{-}2)$$

また、パターンの順番はランダムであるが、必

---

Probablistic Analysis of Weight Vector in Back Propagation Learning
Masahiko AIZAWA*, Tomoko YAMAZAKI*, Sadamu OHTERU* and Shuji HASHIMOTO**
*WASEDA University, **TOHO University

ず２回づつ同一パターンを提示する場合の、A,Bはそれぞれ

$$A_i = -\varepsilon \frac{\partial \langle E \rangle}{\partial w_i} + O(\varepsilon^2) \qquad \underline{i=1,2}$$
$$B_i = 2\varepsilon^2 [\langle (\frac{\partial E}{\partial w_i})^2 \rangle - A_i^2] + O(\varepsilon^2) \qquad (4\text{-}3)$$

となり、平均収束方向は式（４－２）と変わらないが、分散が大きくなり、収束までの時間のばらつきが増大することがわかる。このように学習の確率過程を考えると、パターン提示の方法に対する学習挙動の相違がわかる。

## ５　おわりに

誤差逆伝播学習のポテンシャル関数とランダムなパターン提示の場合の荷重ベクトルの確率過程について定式化した。 続いて、具体例によって、パターンの提示の方法による学習の挙動の変化についても検討した。現在、他の学習方式についても同様の考察を行っているが、忘却効果を導入したり、出力関数を変更することにより学習の効率が向上すること等も判明した。また学習過程で重要な平均学習時間も、荷重ベクトルが最初にポテンシャルの底に達するまでの平均時間とみなすことができるので、本文で定式化した確率過程の初期通過問題に帰着できる。

＜　参考文献　＞


1)甘利：”学習識別の理論”，電子通信学会誌,vol.50,pp.1272-1279,(1967)

2)D.E.Rumelhart,J.L.McClelland and the PDP Research Group:"Parallel Distributed Processing",vol.1&2,The MIT Press ,(1986)

3)橋本、大照、加藤：”ストカスティック計算機の動作解析”、電子通信学会論文誌、59-D,11,p.754(1970)


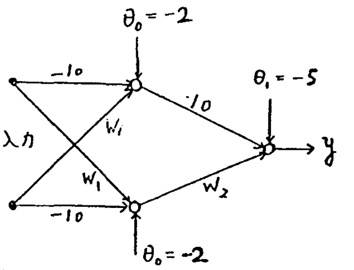


図１　２層３素子の対称ネット

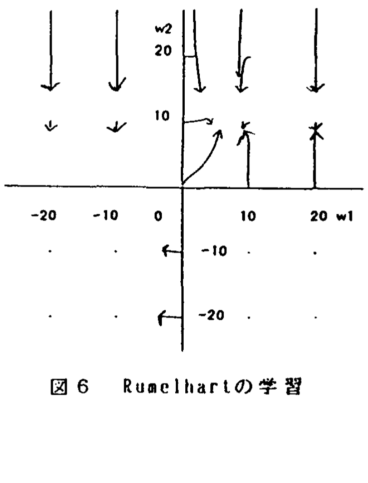


図６　Rumelhartの学習

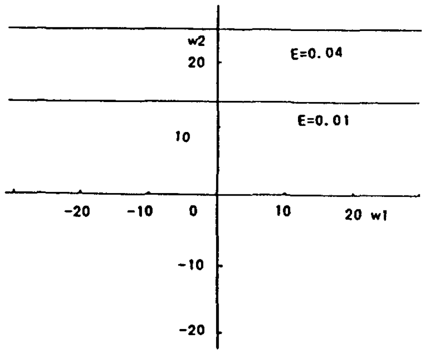


図２ 排他的論理和のポテンシャル関数E
パターン(0,0)

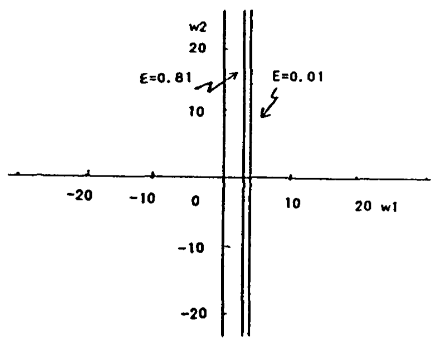


図３　パターン(0,1)

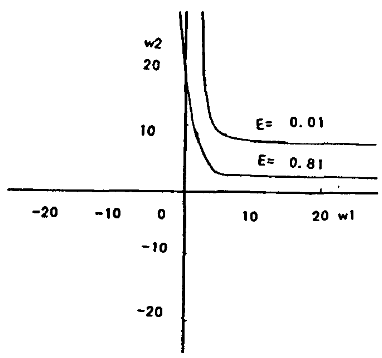


図４　パターン(1,0)

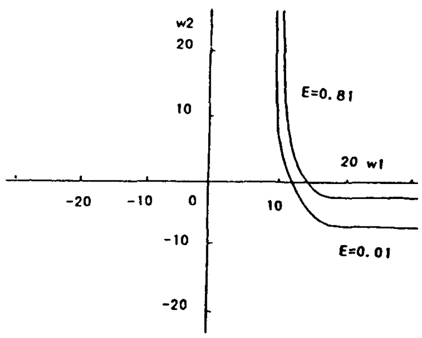


図５　パターン(1,1)