バックプロパゲーション法の学習過程の可視化

4H-10

川村旭　渡部信雄　木本隆(富士通研究所)

1. はじめに

本研究では、ニューロ・コンピュータの学習規則の一つであり、広く用いられているバックプロパゲーション法の学習過程の可視化による調査手段を提案する。また、適用例として、0－1問題のXORの学習を取り上げる。

2. バックプロパゲーション法の説明

バックプロパゲーション法で用いられるニューラル・ネットワークの構造は、図1に示す通り、層状に配列されたユニット(●で示す，①は常に1を出力するユニットであり閾値を重みで実現するのに用いる)と、重みw_{ij}を持つ層間結合(──で示す)とから成る。ユニットでは以下に定式化される通り下層からの入力の加重和をとった後シグモイド関数による閾値処理を行い上層への出力とする。

$$u_{pj}=1/(1+\exp(-\Sigma u_{pi}w_{ij}))$$

ここで、u_{pi}はパターンpのときのユニットiの出力値。

学習は以下の通り誤差を定義し、それを最小化する重みを探索することにより行う。

$$E=\Sigma_p\Sigma_j(u_{pj}-d_{pj})^2$$

ここで、d_{pj}はパターンpのときの出力ユニットjの出力に対応する教師信号である。

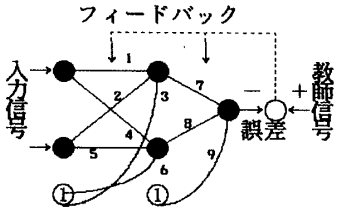
探索の為の重み更新規則は、以下の通りである。

$$\Delta w_{ij}(t)=-\varepsilon\frac{\partial E}{\partial w_{ij}(t)}+\alpha\Delta w_{ij}(t-1)$$

ここで、ε，αは制御パラメータである。

3. 可視化による調査法の説明

学習過程の調査法は、まず多次元の重み空間における、2次元の切断平面上の誤差分布の表示を行う。その切断平面は、最適重みの位置ベクトルと、1つ選択した座標軸とを含むものとする。その平面上の任意の点の位置ベクトルw_{sur}は次の通り指定される。



XOR

入力		教師
0	0	0
0	1	1
1	0	1
1	1	0

図1．ニューラル・ネットワークの構造と学習パターン

$$w_{sur}=w_{axis}e_{axis}+w_{opt}e_{opt}$$

ここで、e_{axis}は、選択した座標軸を示す単位ベクトル、e_{opt}は最適重みの位置ベクトルをe_{axis}に対して正規直交化したものである。w_{axis}，w_{opt}は任意の実数である。その表示により、重み空間に於ける誤差の分布の特徴が判る。また、その誤差分布の表示上に重みの探索過程を重ねて表示を行うことにより、重み更新規則の良否を視覚的に調査することができる。

4. 適用例

適用例として、XORを学習する場合を取り上げる。ニューラル・ネットワークの構造と学習パターンは図1に示す通りである。結合の線の付近の数字は各重みにつけた番号である。図2に、各重みの座標軸を選択した場合の誤差分布の表示を示す。なおw_1とw_2、w_4とw_5、w_7とw_8に関しては同じ分布であるので表示を共用した。

この図に於いて特徴的なことは、重み空間の原点から放射状に伸びる深い谷が存在することである。この谷の軸線は、最適重みの位置ベクトルに沿っている。また、原点から谷に沿って行くと、誤差が急に小さくなるところがあり、その後は誤差は緩やかに小さくなり、無限遠で0となる。また原点から遠ざかるにつれて、誤差がほとんど0の領域が拡がるとともに、谷の両側の壁の勾配が大きくなっていく。先に示したバックプロパゲーション法の重み更新規則は、これらの性質をうまく利用している。

なお、Hintonの文献[1]に、重み空間に於ける誤差の分布を表示し、重み探索過程を観察する方法が記載されているが、それは重みの現在位置中心の局所的な表示である。

5. まとめ

本研究で提案する学習過程の可視化法をXOR学習に適用した。その結果、この種の0－1問題の場合の誤差分布の特徴である峡谷状の形の大局的な分布を表示することができた。今後は、この表示上に探索過程を重ねて、バックプロパゲーション法のパラメータ調節法、及び重み更新規則自体の調査・改良を行っていく。

〔参考文献〕

1) D.C.Plaut and G.E.Hinton,"Learning sets of filters using back-propagation",Computer Speech and Language(1987)2,35-61

Visualization of back-propagation learning processes
Akira KAWAMURA,Nobuo WATANABE,Takashi KIMOTO
Fujitsu Laboratories Ltd.

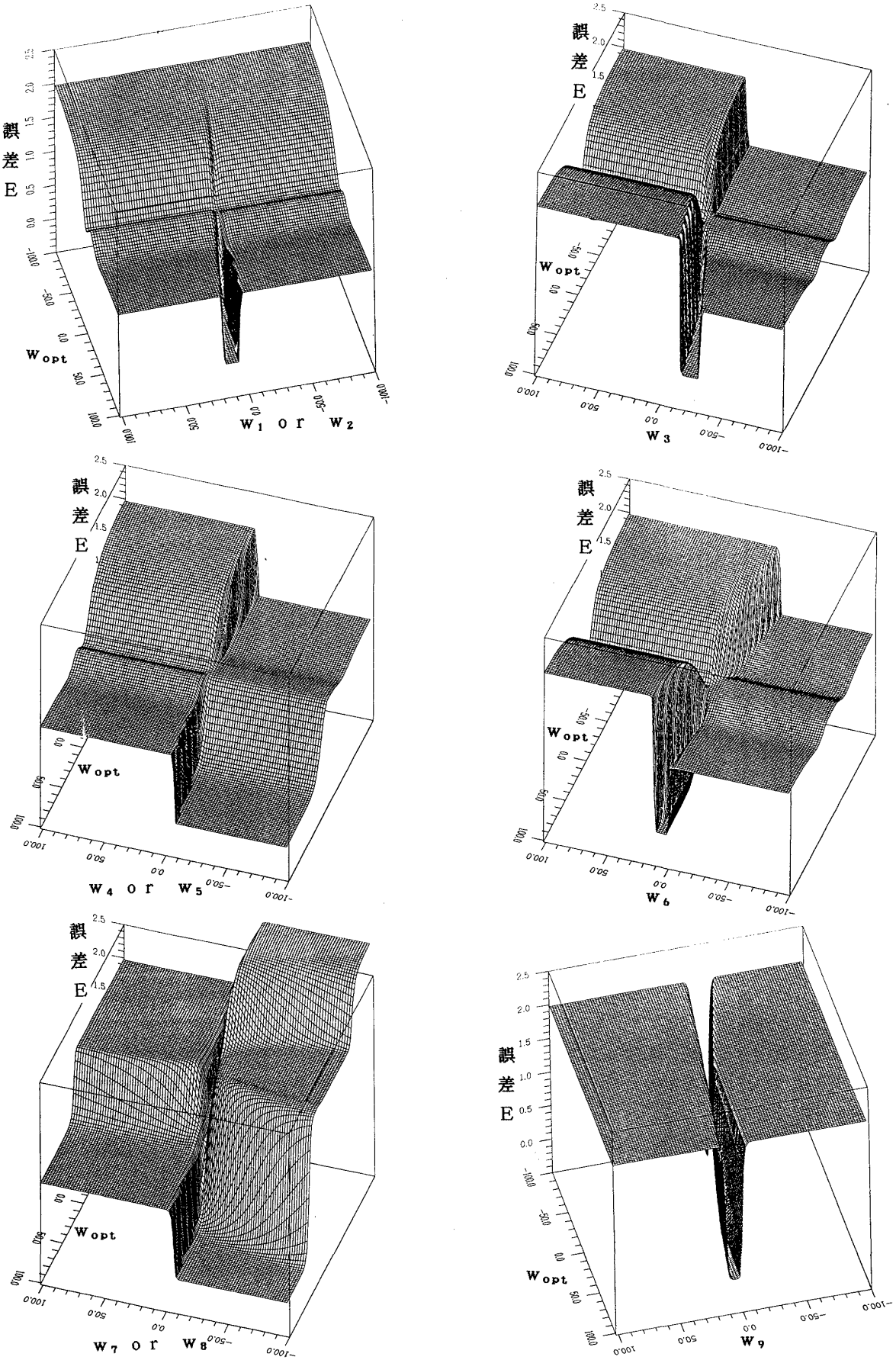


図2．XOR学習の場合の誤差分布の表示