3C－3

並列順伝播学習方式

阿部重夫

（株）日立製作所　日立研究所

1.　はじめに

多層のニューラルネットの学習方式として、教師データを一個ずつ見ながら誤差が最小になるように結合の重みを変えていく逆伝播法が広く用いられている。しかしながらこの方法では前に学習したパターンの情報を保持した上で重みを変えていくことが難しいため、一般に収束が遅く、単純な排他的論理和を収束させるためにも相当なパラメータのチューニングを必要としている。本論文では上記の問題点を解決するため教師パターンを全てみて順方向に重みを変えていく並列伝播学習方式の提案を行い、排他的論理和等で収束特性の評価を行う。

2．ニューラルネットの定義

3層から4層以上のニューラルネットでの学習方式に拡張することは容易に行えるからここでは図1の3層ニューラルネットに限定して考える。

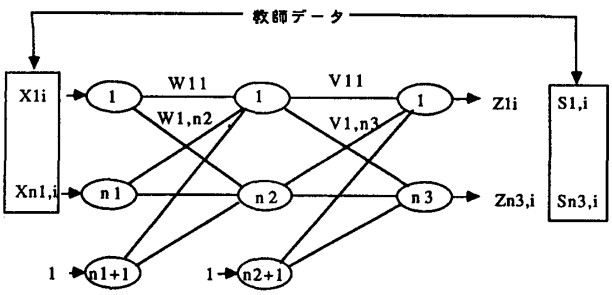

図1 3層ニューラルネット

図において

$$Z_{il}=f_i(U_{il})=1/(1+\exp(-U_{il}/T))$$
$$U_{il}=\sum_{l} V_{ji}Y_{jl} \qquad (1)$$
$$Y_{jl}=G_j(Y_{jl})=1/(1+\exp(-T_{jl}/T))$$
$$T_{jl}=\sum_{k} W_{kj}X_{kl}$$

である。学習とは（1）式のVji，Wkjを次の誤差が最小となるように決めることである。

$$E=1/2\sum_{l}\sum_{i}(Z_{il}-S_{il})^2 \qquad (2)$$

3．学習の考え方

（2）式を最小化するかわりに逆伝播法では[1]、一つの教師データに対して

$$E=1/2(Z_{il}-S_{il})^2 \qquad (3)$$

が最小となるように出力層の重みVjiを変え、ついで入力層の重みWkjを変える。そしてこれを全ての教師データについて繰り返して行くことにより解を求めるものである。　この方法の問題点は、教師データを一つずつ処理していくことにより前に覚えたデータを忘れるため学習の効率が極めて悪いことである。また学習のための加速パラメータを応用ごとにチューニングしなくてはならないというわずらわしさがある。いくつかの改良方式が提案されているが、それらの多くは教師データを一個ずつ処理することは変えておらず本質的な改善になっていない。

上記問題を解決するには、（3）式を逐次的に最小化するのではなく（2）式をそのまま最小化することが必要である。それには（2）式をVji,Wkjで偏微分した式を0と置いてそれを解けばよいと思える。即ち

$$\partial E/\partial V_{ji}=0$$
$$\partial E/\partial W_{kj}=0$$

をNewton-Raphson法で解くことである。

Learning by Parallel Forward Propagation
Shigeo ABE
Hitachi, Ltd.

しかしながらこの方法では解が求まらない。それは各層の重みが変数であるため、ある点が（２）式を極小化する場合、Wkj,Vjiをある方向に微小変化した点も極小値を取りうるからである。即ち（２）式を極小化するWkj,Vjiは点ではなく連続した曲線となり、上記の方法では最小化する方向が定まらない訳である。

問題は変数Wkj,Vjiが直列につながっていることにあるから、逆伝播法と同じように一層ごとに重みを決めてゆけばNeuton-Raphson法の適用が可能となる。入力側からでも出力側からでも重みを決めてゆくことができるが、逆伝播法と区別するために並列順伝播法とよぶことにする。

４． アルゴリズム

４．１ Wkjの決定

Vjiを一定として,（２）式の最小化を図る。ここでfiは一価関数であるから、Silを出力する入力Ailは

$$A_{il} = -T\log(1/S_{il}-1) \tag{3}$$

である。これより(2)式の替わりに、

$$E' = 1/2\sum_{l}\sum_{i}\{\sum_{j} V_{ji}g_j(\sum_{k} W_{kj}X_{kl})-A_{il}\}^2 \tag{4}$$

を最小化することにすれば、収束性を高めることができる。（４）式が最小となる条件は、

$$\partial E'/\partial W_{kj} = \sum_{l}\sum_{i}\{\sum_{j} V_{ji}g_j(\sum_{m} W_{mj}X_{ml})-A_{il}\} \times \partial g_j/\partial U_{jl}X_{kl} = 0 \tag{5}$$

である。

４．２ Vjiの決定

今度はWkjを一定として（２）式の最小化を図る。即ち

$$E' = 1/2\sum_{l}\sum_{i}\{\sum_{j} V_{ji}Y_{jl}-A_{il}\}^2 \tag{6}$$

をYjl一定として求めればよいが、これは一度の計算で求まる。

５． 評価

排他的論理和(EOR)及び文献２の入力が偶数個あるいは全て０のとき１を出力する偶数判定回路で収束特性を評価した。このとき各々の回路に対するニューロン数は入力側から2、2、1 及び7、2、1とした。初期値は

$$V_{ji}, W_{kj} = \mathrm{Bias} + \alpha\,\mathrm{Rand}$$

で与えた。但しRandは-0.5から0.5の範囲の一様乱数とした。また出力値は0.1,0.9として収束判定はおのおの0.00001および0.00005として100回試行した。その結果を表1及び2に示す。どちらの場合も0.5付近に初期値を与えたとき良好な結果が得られた。ただ収束したVjiが大きな値になる場合が有り、この点に関しては更に検討が必要である。

表1　EORの収束特性(100回試行)

Bias	α	収束回数	平均ステップ数
0.5	0.4	88	5.0
0.5	0.1	99	2.6
0.5	0.01	98	2.5
0.5	0.001	98	2.5

表2　偶数判定回路の収束特性(100回試行)

Bias	α	収束回数	平均ステップ数
0.5	0.1	69	26
0.5	0.01	86	14
0.5	0.001	85	12

6.おわりに

教師データを全てみて階層ごとに最適な方向に重みを修正する並列順伝播方式の提案を行った。また排他的論理和などの簡単な回路で収束特性の評価を行い良好な結果を得た。

参考文献

１． D.E.Rumelhart et al,"Parallel Distributed Processing",Vol. 1,2,MIT Press, Cambridge,Mass

2.N.Baba,"A New Approach for Finding the Global Minimum of Error Function of Neural Networks", Neural Networks,Vol. 2, pp.367-373,1989