5F-1

Rumelhart型ネットワークに関する基礎研究
全パターン識別のための学習法

阿部 紳聡 、 嘉数 侑昇

北海道大学

1. 緒言

Rumelhart型ネットワークによる学習に関して、パターンの識別率の向上は1つの課題である。
しかし、 Rumelhartの提唱したエラー関数では、学習によって識別率は向上するが、それは全パターンを識別することとは本質的に異なる。
本研究では、全パターンを識別するための新しいエラー関数を導入し、それを使った学習法を明らかにする。また、計算機実験によってその有効性を検証する。

2. エラー関数Eの問題点

Rumelhartの提唱したエラー関数Eは次式で表される(図1参照)。

$$E=\frac{1}{2}\sum_{c}\sum_{j}(y_{j,c}-d_{j,c})^2 \quad ①$$

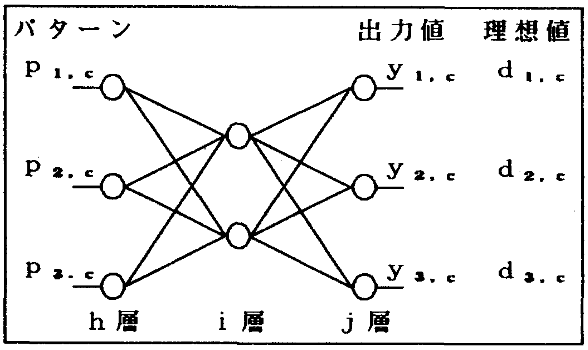

図1. Rumelhart型ネットワーク

出力値の安定性から、Eによる学習で、ローカルミニマに収束したとき、正しく識別されない出力$y_{j,c}$ の多くは次式を満たす($d_{j,c}=0$または1の場合)。

$$|y_{j,c}-d_{j,c}|\simeq 1 \quad ②$$

Eに対する出力層の重みw_{ji}の変化量は、

$$\frac{\partial E}{\partial w_{ji}}=(y_j-d_j)y_j(1-y_j)y_i \quad (c:fix) \quad ③$$

で与えられ、$d_j=0$または1のとき

$$\left.\begin{aligned} d_{j,c}=0 \text{かつ} y_j=2/3 &\Rightarrow \frac{\partial E}{\partial w_{ji}}=4/27y_i \\ d_{j,c}=1 \text{かつ} y_j=1/3 &\Rightarrow \frac{\partial E}{\partial w_{ji}}=-4/27y_i \end{aligned}\right] \quad ④$$

でそれぞれ大きさが最大となる。しかし、式②の場合、

$$\left.\begin{aligned} d_{j,c}\simeq 0 &\Rightarrow \frac{\partial E}{\partial w_{ji}}\fallingdotseq -y_j y_i \\ d_{j,c}\simeq 0 &\Rightarrow \frac{\partial E}{\partial w_{ji}}\fallingdotseq -y_j y_i \end{aligned}\right] \quad ⑤$$

となって、w_{ji}の変化量は式④に比してはるかに小さくなるため、y_j の修正が難しくなる。

3. エラー関数の分離

エラー関数Eは各パターンにおけるエラー関数値Ecの和でも表される。

$$E=\sum_{c}E_c=\sum_{c}\frac{1}{2}\sum_{j}(y_{j,c}-d_{j,c}) \quad ⑥$$

式⑥を2つに分離し、それらをE1, E2とする。このとき、E1, E2を次式のようにとる。

$$\left.\begin{aligned} E_1 &: |y_{j,c}-d_{j,c}|<e \text{ なる } E_c \text{の和} \\ E_2 &: |y_{j,c}-d_{j,c}|\geqq e \text{ なる } E_c \text{の和} \end{aligned}\right] \quad ⑦$$

$$E=E_1+E_2 \quad ⑧$$

eは出力値と理想値との許容誤差で、以下の範囲である。

$$0<e\leqq 1 \quad ⑨$$

最急降下法による学習則を適用すると、E, E1およびE2 による重み変化量$\delta w, \delta w_1, \delta w_2$は次式となる。

$$\left.\begin{aligned} \delta w &= -\varepsilon\frac{\partial E}{\partial w} \\ \delta w_1 &= -\varepsilon\frac{\partial E_1}{\partial w} \\ \delta w_2 &= -\varepsilon\frac{\partial E_2}{\partial w} \end{aligned}\right] \quad ⑩$$

εは重み変化の幅を表し、$\varepsilon>0$である。また、

$$\delta w=\delta w_1+\delta w_2 \quad ⑪$$

δwとδw_2 の方向が近いすなわち、

$$\cos^{-1}\frac{w\cdot w_2}{|w||w_2|}<\beta \quad ⑫$$

であれば、Eによる学習によりE2 の値は減少するが、δwとがδw_1 に近い方向ならばE2 の値が増加し、出力値と理想値との誤差が許容誤差eよりも益々大きくなる。

ここで、E2 のみによる学習を行うことにより、許容誤差の範囲内に収まるパターンの数が増える(図2)。

Study on Rumelhart's network.
Nobuaki Abe, Yukinori Kakazu
Hokkaido University

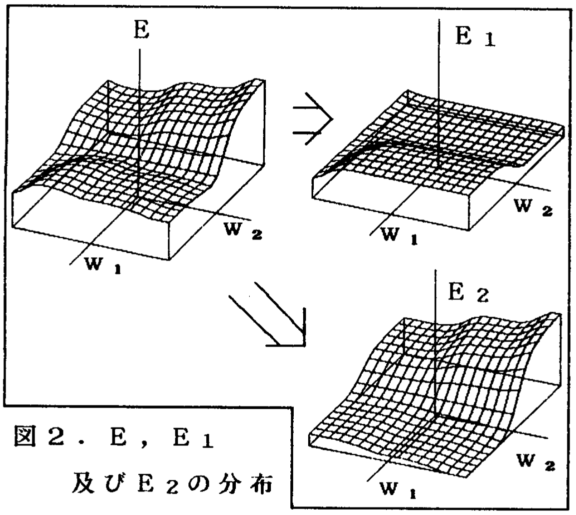

図２．Ｅ，Ｅ１
及びＥ２の分布

４．新しいエラー関数

式①のエラー関数に替えて、今回提案する新しいエラー関数E'は次式のように表される。

$$E = \frac{1}{2}\sum_{c}\sum_{j} def_{j,c} \qquad ⑬$$

ここでｄｅｆj,c は、以下のように定める。

$$def_{j,c} = \begin{cases} 0 & : |y_{j,c}-d_{j,c}| < e \text{ 時} \\ (y_{j,c}-d_{j,c})^2 - e^2 & : |y_{j,c}-d_{j,c}| \geqq e \text{ 時} \end{cases} \qquad ⑭$$

理想値ｄj,c が0または1の2値をとるとき、出力ｙj,c が以下であれば「出力値はｄj,c である。」と識別できる。

$$| y_{j,c} - d_{j,c} | < 0.5 \qquad ⑮$$

すなわち、ｅ≦０．５のとき、全ての出力パターンが正しく識別されたならばE'＝0となる。

エラー関数をE'とすることにより、許容誤差ｅ内の出力に関しての学習は行われず、ｅ外の出力のみの学習が行われる。これは、3節のE2 による学習と等価である。

５．E'とｅによる学習法

２節で述べたように、出力値と理想値との誤差が１に近づかないように学習することが望ましいので、許容誤差ｅの値は最初大きい方がよい。以上から、E'とｅによる学習法として以下のプロセスを考える。

① ｅ＝ｅo （≃１）とする。
② E'について学習を行う。
③ E'＝0になったら、
　　ｅ←ｅ－γ　（γ＞0）に更新する。
④ ②、③を繰り返す。

ｅ＝０．５のとき、全パターン識別が可能となるが、出力が不安定なので、ｅ＝０．１～０．２程度で学習を終了する。（ｄj,c= 0または1のとき）。

６．実験

簡単なネットワークを組んで計算機実験を行った。３層のRumelhart 型ネットワークでニューロン数は入力層１０、出力層１０である。パターン数は８１個で、隠れ層が７、１０、１３、１６個のときEおよびE'（E2)による学習を行った（図３，４）。

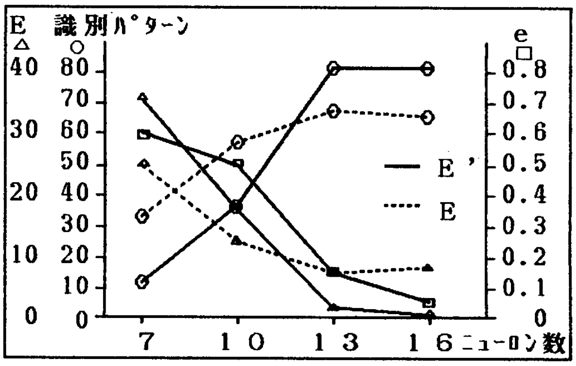

図３．　ＥとＥ'による学習

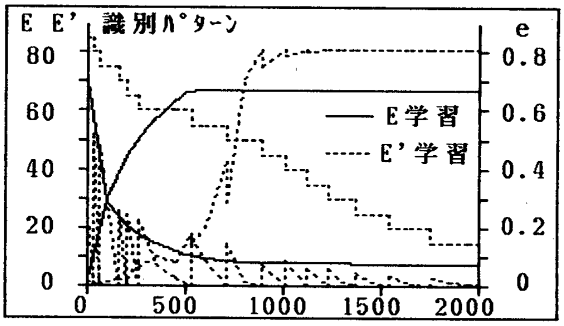

図４．学習によるパラメータの変化

図3より、Eによる学習では隠れ層のニューロン数を増やしても全パターンが識別されなかったが、E'による学習では、ニューロン数１０より多いとき識別された。

また図4から、E'による学習の方が初め識別パターン数が少ないが、ｅ≃０．２５くらいで識別パターン数が安定（全パターン）する事が確かめられた。

７．考察

中間層のニューロン数が少ないときｅの値は０．５より下がらなかった。Eによる学習でも識別率は低いので、このときのネットワークはパターンを識別するのに十分な大きさがないと考えられる。

ｅの値を見ることによって入出力パターンに対するネットワークの規模がわかるので、ｅを使うことによってネットワークを動的に拡張することができると思う。

８．結言

新しいエラー関数を導入して、全パターンを識別するための学習法を明確化した。

計算機実験を行い、学習法の有効性を確かめた。

参考文献

1）阿部紳聡 「動的Rumelhart型ネットワークに関する基礎研究」
　電気関係学会北海道大会講演集 p292 (1988)
2）E.Rumelhart et al. "Learning representations by back-propagating errors"
　NATURE VOL.323 9 OCT. 1986
3）甘利俊一 「神経回路網の数理」 産業図書(1978)