

ニューラルネットワークの耐最悪故障化学習

西垣 正勝[†] 都筑 輝泰^{††} 曾我 正和^{†††}

A Learning Algorithm Considering the Tolerances of the Worst-Case Fault of Neural Networks

Masakatsu NISHIGAKI[†], Teruyasu TSUZUKI^{††}, and Masakazu SOGA^{†††}

あらまし ニューラルネットワークを利用することにより、フォールトトレラント回路を効率良く実現することを目的とする。故障の範囲としては、中間層-出力層間の単一断線故障を想定する。本論文では、通常の誤差逆伝搬アルゴリズムを用いて各学習セットに対する最悪故障時の出力を教師信号に一致させることにより耐故障性を獲得する「耐最悪故障化学習」アルゴリズムを提案する。シミュレーションを通じ、本学習方式の学習効率、及び、本学習方式により得られる耐故障ニューラルネットワークの耐故障強度を評価し、良好な結果を得た。また、本学習方式がニューラルネットワークの汎化能力をも向上させることについても報告する。

キーワード 耐故障性, 最悪故障, 汎化能力, 階層型ニューラルネットワーク, 誤差逆伝搬学習

1. ま え が き

本論文では、ニューラルネットワークによるフォールトトレラント回路の実現を目指す。ニューロンは入力荷重がしきい値より大きいと出力が決定されるため、しきい値をまたいで超えるような故障でない限り、内部がある程度壊れても同じ出力を保つことができると考えられる。つまり、ニューラルネットワークはフォールトトレラント回路を作るに適した回路であると考えられる。

回路に耐故障性をもたせるためには二つの「冗長性」が必要となる。

1. 情報の冗長性：故障部位の情報を正常部位の情報から復元できなければならない。

2. 構造の冗長性：冗長を含んだ情報を記憶するために構造的（回路規模）にも冗長性が必要となる。

すなわち、回路に冗長成分（冗長素子）を与え、それらに情報（回路の入出力関係）を適切に分散させても

たせることにより耐故障性が実現されるといえる [1].

しかし、階層型ニューラルネットワークにおける故障を考えたとき、冗長ユニット（冗長ニューロン）をもつネットワークを通常の誤差逆伝搬アルゴリズムで学習させても、すべてのユニットが有効に活用されるとは限らない。誤差逆伝搬アルゴリズムでは正常時の誤差を極小にするようにパラメータを修正するだけであるため、故障を考慮した学習は行われず、情報が全ユニットに適切に分散化される保証（故障部位の情報を正常部位の情報から復元できるように情報を冗長化しながら自己組織化を行う能力）はない。

これに対し、耐故障性を有するニューラルネットワークを作る学習アルゴリズムが提案されている。正常状態だけでなく全故障状態を想定しての学習を行う FTBP 学習アルゴリズム [2] や、最悪故障の原因となる結合荷重の大きさを制限する学習法 [3] などが挙げられる。しかし、前者にはすべての故障状態を学習するため時間がかかるという欠点が存在し、また、後者は断線による悪影響を減らすために結合荷重を小さくすることしか行わず、耐故障化に対する指針に少々疑問が残る。

本論文では、最悪故障状態を予測して最悪故障時の出力のみを教師信号に一致させる新しいタイプの耐故障化学習アルゴリズムを提案する。更に、提案する耐故障化学習がニューラルネットワークの汎化能力を向

[†] 静岡大学情報学部情報科学科, 浜松市
Department of Computer Science, Faculty of Information,
Shizuoka University, Hamamatsu-shi, 432-8011 Japan

^{††} シャープ株式会社 IC 事業本部, 天理市
Sharp Corporation, Integrated Circuits Group, Tenri-shi,
632-8567 Japan

^{†††} 岩手県立大学ソフトウェア情報学部, 岩手県
Faculty of Software and Information Science, Iwate Prefec-
tural University, Takizawa-mura, Iwate-ken, 020-0173 Japan

上させる性質をも有するというシミュレーション結果が得られたことを報告する。

なお、本論文では故障を中間層-出力層間の結合において発生する単一断線と仮定し、ここでは中間層-出力層間の単一断線故障に対する耐故障性を与えるアルゴリズムについて述べる。したがって、最悪故障とは中間層-出力層間の単一断線故障の中で最大の出力誤差を引き起こす故障を指す。本アルゴリズムは多層ニューラルネットワークの全結合の断線に対して拡張可能であるが、本論文では故障集合を簡素にし、アルゴリズムをわかりやすく示すことに努める。なお、ニューロン間の結合が大部分を占めるニューラルネットワークにおいては、故障を断線に絞るという仮定は妥当であるといえる [2], [3]。

2. 記 述

本論文で仕様する記号を記しておく。また、図1にネットワークの概略図を示す。

[記述1] 第 p パターンをネットワークの入力層に入

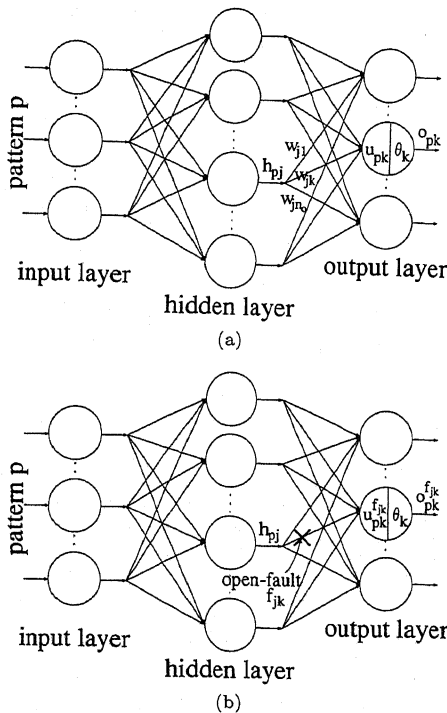


図1 (a) ニューラルネットワーク
(b) 故障時のネットワーク
Fig. 1 (a) Neural network.
(b) Faulty network.

力した際の、中間層第 j ニューロンの出力を h_{pj} 、出力層第 k ニューロンの出力及び教師信号を o_{pk} 、 t_{pk} とする。

[記述2] 中間層第 j ニューロン-出力層第 k ニューロン間の結合枝を l_{jk} とし、その結合荷重を w_{jk} とする。

[記述3] 断線故障を f とする。特に、結合枝 l_{jk} に断線故障が起こった状態を f_{jk} と記す。

[記述4] 故障 f が起こったネットワークにおいて、第 p パターンを入力層に入力した際の出力層第 k ニューロンの出力を故障出力 o_{pk}^f とする。特に、故障 f_{jk} が起こった場合の故障出力を $o_{pk}^{f_{jk}}$ と記す。

[記述5] 第 p パターンを入力した際の出力層の第 k ニューロンの出力 o_{pk} ($k = 1, 2, \dots, n_o$) は

$$o_{pk} = g(u_{pk}) \tag{1}$$

$$u_{pk} = \sum_{j=1}^{n_h} w_{jk} \cdot h_{pj} + \theta_k \tag{2}$$

で表される。ここで、 n_h 、 n_o は中間層及び出力層のニューロン数、 u_{pk} 、 θ_k は出力層第 k ニューロンの入力の荷重和及びしきい値、 $g(x)$ はシグモイド関数

$$g(x) = \text{sigm}(x) = \frac{1}{1 + \exp(-x)} \tag{3}$$

である。

[記述6] 故障 f が生じた際の u_{pk} を u_{pk}^f と記す。特に、故障 f_{jk} が起こった場合の u_{pk} を $u_{pk}^{f_{jk}}$ と記す。故障 f_{jk} により h_{pj} の入力がか断されるので

$$u_{pk}^{f_{jk}} = u_{pk} - w_{jk} \cdot h_{pj} \tag{4}$$

である。

3. フォールトトレラント学習アルゴリズム

ここでは、丹ら [2] 及び Hammadi ら [3] によって示された耐故障化学習アルゴリズムについて説明する。なお、これらのアルゴリズムはネットワーク中の全結合の単一断線故障に対する耐故障性を与えるものである。

3.1 FTBP 学習

通常の学習は、誤差

$$E = \sum_p \sum_k (t_{pk} - o_{pk})^2 \tag{5}$$

を考え、 $E < \epsilon$ を目的として行われる。ここで ϵ は許容誤差である。これに対し、丹らは

$$E_{FT} = \sum_f \sum_p \sum_k (t_{pk} - o_{pk}^f)^2 \quad (6)$$

という評価関数を導入した。 $E_{FT} < \epsilon$ が満たされるならば、故障出力も教師信号に一致するので、耐故障性を学習したことになる。ここで、 f は考えられ得るすべての故障状態であり、正常状態（故障なしという故障状態）もここに含まれる。 $(t_{pk} - o_{pk}^f)^2$ は常に正であるので $E \leq E_{FT}$ であり、 $E_{FT} < \epsilon$ が満たされれば、 $E < \epsilon$ も満たされる。したがって、通常の誤差逆伝搬学習において、評価関数を E から E_{FT} へと変更するだけで耐故障化学習が可能となる。丹らはこれを FTBP (Fault-Tolerant Back Propagation) 学習と呼んだ。FTBP 学習により得られた回路は耐故障性を有するだけでなく、その汎化能力に対しても向上が認められた。

しかし、FTBP 学習ではすべての故障状態を学習セットとして与える必要があり、学習に時間がかかりすぎる。

3.2 シナプス結合荷重の Relevance を利用した学習

このアルゴリズムは、断線故障時に出力誤差を生じさせるシナプスの結合 w_{jk} に注目する。結合 l_{jk} が断線したときの影響度 (Relevance) を以下の式によって求める。

$$R(l_{jk}) = \max_{p,k} |o_{pk} - o_{pk}^{f_{jk}}| \quad (7)$$

式 (7) は、結合枝 l_{jk} (結合荷重 w_{jk}) の断線に対する影響度 $R(l_{jk})$ を、出力に現れる誤りの最大値として定義している^(注1)。そして、すべての l_{jk} に対する $R(l_{jk})$ を評価し、その中で最大の R を与える l_{jk} が、断線時に最大の誤りを回路に生じさせる結合枝である。この結合枝の結合荷重に

$$w_{jk} \rightarrow \frac{w_{jk}}{1 + \lambda R(l_{jk})} \quad (8)$$

で示される修正を施すことで、その結合の断線が引き起こす出力の誤りの大きさ (断線による悪影響) を小さくすることができる。ここで、 λ はペナルティ係数と呼ばれる正の定数である。この操作を続けることにより、最終的に断線による悪影響が現れることのない回路 (フォールトトレラント回路) が実現される。

しかし、文献 [2] では、EX-OR 回路の耐故障化のために、中間層-出力層間の結合荷重が十分に大きくなることが報告されており、結合荷重を小さく制限することしか行わない Hammadi らの学習法の指針に少々疑問が残る (詳しくは付録参照)。また、 R を求めるためにすべての結合の断線に対して $o_{pk}^{f_{jk}}$ を計算する必要があり、かなりの手間がかかる。文献 [3] では、シグモイド関数をテイラー展開し、その 2 次項までを用いて近似することにより計算を高速化している。しかし、その計算はなお複雑であり、かつ、大きな結合荷重をもつ結線の断線に対しては $o_{pk}^{f_{jk}}$ を近似する能力が低い [4]。

4. 耐最悪故障化学習

4.1 耐最悪故障化学習法

本論文では、正常状態の入出力関係のみを学習セットとし、各入力パターンに対する最悪故障状態の出力を予測しながら、最悪故障出力を教師信号に近づけることにより耐故障性を得るといった新しいタイプの学習アルゴリズムを提案する。

本アルゴリズムは

- 各パターンに対する最悪故障状態の学習を行うだけであるから、FTBP 学習 [2] のように学習セットが増加することはない、

- 最悪故障出力を教師信号に近づけるために通常の誤差逆伝搬学習を使用する (Relevance を利用した学習 [3] のように結合荷重の大きさに制限を与えるタイプのアルゴリズムではない) ので、適切に結合荷重が最適化されることが期待できる

という長所を有する。

本学習方式においても、最悪故障出力を評価するためにすべての $o_{pk}^{f_{jk}}$ を計算する必要性は残る。本論文では、シグモイド関数の飽和特性・単調増加性を利用することにより、計算の高速化を図る (5.2 参照)。

本論文では、最悪故障に対する耐故障化を行う本学習方式を耐最悪故障化学習と呼ぶ。

4.2 耐最悪故障化学習アルゴリズム

耐最悪故障化学習では、教師信号に対し最大の誤差を与える故障状態のみを考慮した学習を行う。これにより全故障状態を学習データとして与えずとも耐故障回路を獲得することができる。

ここで最悪故障に対して定義を与える。

(注1): 文献 [3] では、 $R(l_{jk})$ ではなく $R(w_{jk})$ と記述している。

[定義1] 第 p パターンの入力に対して、結合枝 l_{JK} の断線が教師信号に対して最も大きく異なる故障出力を呈するとき、 l_{JK} の断線をパターン p に対する最悪故障と定義する。すなわち、パターン p に対する最悪故障は

$$\max_{f,k} |t_{pk} - o_{pk}^f|$$

を引き起こす故障 f である。ここで、出力層内の全ニューロンの出力に現れる誤りの総和が比較されるのではなく、各ニューロンの各々の誤りが比較され、最悪故障が決定されることに注意する。つまり、出力が最も大きく誤るニューロンが一つでもあれば、他のニューロンの誤りが小さくとも、その状態を引き起こす故障が最悪故障と定義される。

[定義2] パターン p に対する最悪故障が生じた状態 (l_{JK} が断線した状態) をパターン p に対する最悪故障状態 (または単に最悪故障状態) と呼ぶ。なお、故障集合には故障なしという状態も含まれる。

[定義3] パターン p に対する最悪故障状態における出力をパターン p に対する最悪故障出力 (または単に最悪故障出力) と呼ぶ。

[記述7] パターン p に対する最悪故障を f_p^{\max} と記す ($f_p^{\max} = f_{JK}$ である)。

[記述8] f_p^{\max} が生じた際の出力層第 k ニューロンの最悪故障出力を $o_{pk}^{f_p^{\max}}$ と記す。

[記述9] f_p^{\max} が生じた際の u_{pk} を $u_{pk}^{f_p^{\max}}$ と記す。

なお、 f_p^{\max} の断線が起こる結合枝 l_{JK} は各学習反復ごと、各パターンごとで異なり得ることに注意する。また、本論文では中間層-出力層間の結合の単一断線故障を仮定している。したがって、最悪故障とは中間層-出力層間の単一断線故障の中で最大の出力誤差を引き起こす故障を意味する。

耐最悪故障化学習では、 o_{pk} ではなく、 $o_{pk}^{f_p^{\max}}$ を教師信号 t_{pk} に一致させるように学習を行う (図 2)。この操作は、故障出力 o_{pk}^f の集合すべてを教師信号に一致させることに通ずる。この結果、すべての故障状態の出力が教師信号と一致することになり、耐故障性が保証される。

本アルゴリズムの評価関数は次式となる。

$$E_{ME} = \sum_p \sum_k (t_{pk} - o_{pk}^{f_p^{\max}})^2 < \epsilon \quad (9)$$

したがって、評価関数を E (式 (5)) から E_{ME} に変更するだけで耐最悪故障化学習を行うことができる。

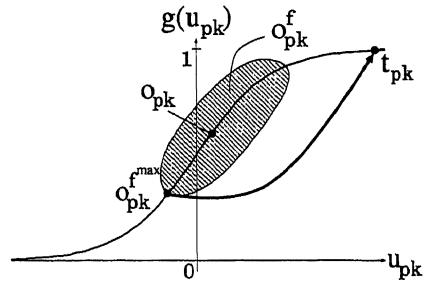


図 2 最悪故障を考慮した学習
Fig. 2 Learning scheme considering the worst-case fault of neural networks.

E_{ME} の極小化には通常の誤差逆伝搬アルゴリズムを用いればよい。

本学習方式の収束性に関する数学的な議論については、まだ解明されていない点も多く、その詳細な検討については本論文の範疇を超える。しかし、最悪故障 f_p^{\max} は各学習反復ごと、各パターンごとに逐次求められ、そのつど、最悪故障を減少させる方向に学習が進められる。したがって、本アルゴリズムにより $E_{ME} < \epsilon$ が成立した時点で、すべての故障に対して耐故障性が満たされることは保証される。なお、故障集合には故障なしという状態も含まれているため、 $E_{ME} < \epsilon$ の成立が $E < \epsilon$ を含むことを明記しておく。

耐最悪故障化学習アルゴリズムを以下に示す。

1. 各パターン p の入力に対して以下を行う。

- (i) 故障状態 (正常状態を含む) の中から最悪故障 f_p^{\max} を引き起こす連結 l_{JK} を見つける。
- (ii) 最悪故障状態の出力 $o_{pk}^{f_p^{\max}}$ を求める。
- (iii) 誤差逆伝搬学習アルゴリズムを用いて、式 (9) の評価関数を極小化する方向に全結合荷重を更新する (なお、 f_p^{\max} を引き起こす結合枝 l_{JK} に対する結合荷重の更新は行わない)。

2. 以上を収束するまで繰り返す。

5. 2 値出力システムの耐故障化

5.1 耐最悪故障化学習による 2 値出力システムの耐故障化

出力として連続値をとるシステムをニューラルネットワークで実現する場合、教師信号は (0 から 1 の) 任意のアナログ値をとる。例として第 P パターンにおける出力層第 K ニューロンに対する教師信号が 0.5 の場合 ($t_{PK} = 0.5$ の場合) を考える。耐故障回路においては、いかなる単一断線故障 f が起こっても o_{PK}^f

が 0.5 とならなければならない。これは、いかなる単一断線故障 f が起こっても $u_{PK}^f = 0$ を保つことと等価である。

一方、出力として 1 または 0 をとる 2 値出力システムの場合、教師信号も 1 または 0 である。例として $t_{PK} = 1$ の場合を考える。出力 1 を呈する領域はシグモイド関数の飽和領域にあたる。したがって、故障出力 $o_{PK}^f = 1$ を満たすためには、 u_{PK}^f が出力 1 の飽和領域に入っていさえすればよい（すべての u_{PK}^f をある一つの値に一致させなくてもよい）。シグモイド関数の飽和特性及び単調増加性から

$$\text{sigm}(a) \approx 1 \text{ ならば,}$$

$$x \geq a \text{ に対して } \text{sigm}(x) \approx 1$$

が成立する。図 2 から $o_{PK}^{f_{\max}}$ に対する $u_{PK}^{f_{\max}}$ が u_{PK}^f の最小値であることがわかるので、 $o_{PK}^{f_{\max}}$ が $t_{PK}(=1)$ に一致しさえすれば、故障出力 o_{PK}^f の集合すべてが教師信号に一致することが保証される。以上は、教師信号が 0 である場合についても同様である。

なお、シグモイド関数の単調増加性により、教師信号が「ある値（アナログ値）より大きい、または小さい」という形で与えられるシステムにおいても同様の性質が成り立つことがわかる。また、出力が 2 値であるシステムであれば、入力ベクトルがアナログ値であっても上述の理論は成立する。

5.2 最悪故障推定法

耐最悪故障化学習では、すべての故障出力 o_{pk}^f を計算し、最悪故障出力 $o_{pk}^{f_{\max}}$ を知る必要がある。ここで、正常状態の u_{pk} や o_{pk} から故障出力 o_{pk}^f や $o_{pk}^{f_{\max}}$ を導出することができれば、最悪故障 $f_p^{f_{\max}}$ を高速に見つけることが可能となる。

耐故障性をもつ 2 値出力システムを耐最悪故障化学習により生成する場合、以下のように、最悪故障 $f_p^{f_{\max}}$ を高速に推定することが可能である。

まず、パターン p に対する最悪故障 $f_p^{f_{\max}}$ の評価を下記の二つのフェーズに分割する。

1. 出力層の各ニューロン k に対して、出力 o_{pk} を最も大きく誤らせる断線故障（これを $f_{pk}^{f_{\max}}$ と記す）を個別に求め、
2. フェーズ 1 で求めた各故障 $\{f_{pk}^{f_{\max}} | k = 1, 2, \dots, n_o\}$ を候補として、その中から最悪故障 $f_p^{f_{\max}}$ を求める。

フェーズ 1 では、シグモイド関数の単調増加性を利用することにより $f_{pk}^{f_{\max}}$ の発見を高速化すること

ができる。出力層第 k ニューロンにつながる結合枝 $l_{jk}(j = 1, 2, \dots, n_h)$ の断線故障 f_{jk} により、ニューロンの出力 o_{pk} は

$$\begin{aligned} o_{pk}^{f_{jk}} &= g(u_{pk}^{f_{jk}}) \\ &= g(u_{pk} - w_{jk} \cdot h_{pj}) \end{aligned} \quad (10)$$

となる。ここで、教師信号が 1 または 0 であること、及び、シグモイド関数が単調増加関数であることから、各 k において

$$\max_j |t_{pk} - o_{pk}^{f_{jk}}|$$

を満たす断線 $f_{pk}^{f_{\max}}$ は

$$\begin{cases} t_{pk} = 1 \text{ ならば } \max_j \{w_{jk} \cdot h_{pj}\} \\ t_{pk} = 0 \text{ ならば } \min_j \{w_{jk} \cdot h_{pj}\} \end{cases}$$

を満たす断線 f_{jk} であることがわかる。ただし、 $t_{pk} = 1$ において $\max_j \{w_{jk} \cdot h_{pj}\} \leq 0$ の場合、及び、 $t_{pk} = 0$ において $\min_j \{w_{jk} \cdot h_{pj}\} \geq 0$ の場合には、故障によってネットワークの出力は教師信号に近づく方向に変位することになるため、故障なし（断線なし）という状態が $f_{pk}^{f_{\max}}$ となる。このように、各 k における $f_{pk}^{f_{\max}}$ の評価には、 $o_{pk}^{f_{jk}}$ を各々計算する必要はなく、 $w_{jk} \cdot h_{pj}$ の大小比較を行えば十分である。

フェーズ 2 においては、各 k ごとに u_{pk} の値が異なるため、 $w_{jk} \cdot h_{pj}$ の大小比較だけでは $\{f_{pk}^{f_{\max}} | k = 1, 2, \dots, n_o\}$ の中から最悪故障 $f_p^{f_{\max}}$ を見つけることはできない。しかし、シグモイド関数ではその単調増加性より

$$a > b \text{ ならば, } \text{sigm}(a) > \text{sigm}(b)$$

が成立するので、 u_{pk}^f が小さいほど $o_{pk}^f = \text{sigm}(u_{pk}^f)$ は 1 から離れ、 u_{pk}^f が大きいほど $o_{pk}^f = \text{sigm}(u_{pk}^f)$ は 0 から離れることがわかる。2 値出力システムでは教師信号は 1 または 0 であるので、フェーズ 1 において求められた $\{f_{pk}^{f_{\max}}\}$ の中から

$$\max_k |t_{pk} - o_{pk}^{f_{pk}^{f_{\max}}}|$$

を満たす最悪故障 $f_p^{f_{\max}}$ を見つけるためには、 $|t_{pk} - o_{pk}^{f_{pk}^{f_{\max}}}|$ の代わりに $u_{pk}^{f_{pk}^{f_{\max}}}$ を比較してやればよいことがわかる。すなわち、

$$\begin{cases} t_{pk} = 1 \text{ ならば } \min_k \{u_{pk}^{f_{pk}^{f_{\max}}}\} \\ t_{pk} = 0 \text{ ならば } \max_k \{u_{pk}^{f_{pk}^{f_{\max}}}\} \end{cases}$$

を満たす断線故障 f_{pk}^{\max} が t_{pk} ($=1$ または 0) から最も大きく誤った出力を引き起こす最悪故障 f_p^{\max} である。ここで、 o_{pk}^{\max} , u_{pk}^{\max} は、故障 f_{pk}^{\max} が起こった際の o_{pk} 及び u_{pk} である。

この結果、 o_{pk}^f の計算をし直すことなく、容易に最悪故障出力を推定することが可能となる。2 値出力システムに対する最悪故障出力の評価を行う本計算法を最悪故障推定法と呼ぶ。最悪故障推定法のアルゴリズムを以下に示す。

- 各学習反復ごと、各パターン p ごとに以下を行う。

1. 出力層の各ニューロン k に対して

$$\begin{cases} t_{pk} = 1 \text{ ならば } \max_j \{w_{jk} \cdot h_{pj}\} \\ t_{pk} = 0 \text{ ならば } \\ \min_j \{w_{jk} \cdot h_{pj}\} = \max_j \{-w_{jk} \cdot h_{pj}\} \end{cases}$$

を満たす結合枝 l_{jk} (結合荷重 w_{jk}) を求める。ここで、各 k に対して J は異なり得る。この断線がパターン p に対する最悪故障の候補 f_{pk}^{\max} となる。

2. $\{f_{pk}^{\max} | k = 1, 2, \dots, n_o\}$ の中から

$$\begin{cases} t_{pk} = 1 \text{ ならば } \\ \min_k \{u_{pk}^{\max}\} = \max_k \{-(u_{pk} - w_{jk} \cdot h_{pj})\} \\ t_{pk} = 0 \text{ ならば } \\ \max_k \{u_{pk}^{\max}\} = \max_k \{u_{pk} - w_{jk} \cdot h_{pj}\} \end{cases}$$

を満たす結合枝 l_{jk} を求める。この断線がパターン p に対する最悪故障 f_p^{\max} となる。

6. 耐最悪故障化学習の評価

シミュレーションを通じて耐最悪故障化学習の評価を行う。耐故障化学習に対する評価の尺度としては、学習効率だけでなく耐故障強度についても考えなければならない。更に、文献[2]では、耐故障化学習がニューラルネットワークの汎化能力をも向上させることが報告されている。したがって本論文では、学習効率、耐故障強度、汎化能力の三つに対する評価を行うこととする。

6.1 学習効率の評価

6.1.1 学習対象

7×7画素のA~Hの八つの文字について文字認識を行うタスクを学習させる。それぞれのピクセルの2値パターンを入力とし、対応する文字の出力のみが1となるように学習を行う(図3)。

ニューロンをしきい値素子と考えると1ニューロンが1ビットに相当する。本タスクの学習においてネットワークの中間層表現の自己組織化が理想的に行われるものと仮定すると、八つのアルファベットを区別するためには中間層は3ユニット(3ビット)必要であることがわかる。そして、出力層では中間層からの3ビットの信号をデコードしているのだと仮定すると、出力層のニューロンはANDゲートのような振舞いをしていると考えられる。一般に単一断線故障をマスクするためには回路を3重化しなければならない[1]が、中間層-出力層の単一断線故障が非対称故障となるように自己組織化が達成できればANDゲートの耐故障化は2重化により実現可能である(図4)。したがって、図3の文字認識を行う耐故障回路を実現するためには、中間層は6ユニット程度以上を要すると予想される。

本節では、入力層49-中間層6-出力層8のネットワークに対して耐最悪故障化学習を行い、図3の耐故障回路を獲得することができるかを調べる。本回路は2値出力の回路であるため、5.2で示した最悪故障推定法が用いられる。

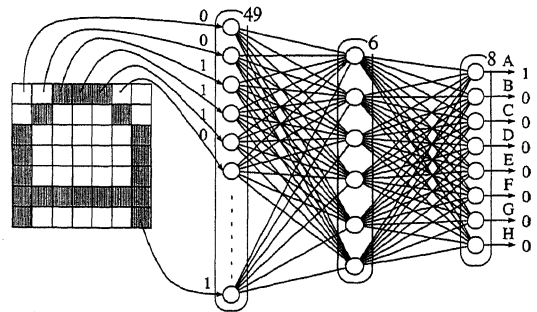


図3 文字認識ネットワーク
Fig.3 8 characters classifier.

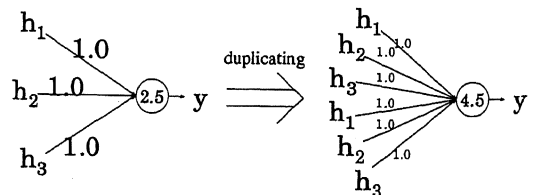


図4 $h_1 \cdot h_2 \cdot h_3$ 回路の耐故障化
Fig.4 Duplication of network to obtain fault tolerance.

表1 シミュレーション結果
Table 1 Simulation results.

耐最悪故障化学習			
ネットワークサイズ	49-6-8	49-8-8	49-12-8
学習成功率	95/100	100/100	100/100
平均学習エポック数	8727.6	1446.8	562.6
平均収束時間 (sec)	12.7	2.7	1.5
1エポック当りの学習時間 (msec)	1.5	1.8	2.7
学習の準収束率	3/100	-/100	-/100
必須結合数 (精度 0.1)	5.67	0.00	0.00
FTBP 学習			
ネットワークサイズ	49-6-8	49-8-8	49-12-8
学習成功率	11/100	91/100	100/100
平均学習エポック数	15295.4	6518.0	775.0
平均収束時間 (sec)	411.6	284.3	68.2
1エポック当りの学習時間 (msec)	26.9	43.6	88.0
学習の準収束率	89/100	9/100	-/100
必須結合数 (精度 0.1)	9.60	1.89	0.00

6.1.2 学習結果

耐最悪故障化学習を行った結果を表1に示す。100回の試行における学習成功率と学習に要したエポック数及び時間を調べた。各試行ごとに、結合荷重としきい値の初期値は $[-5.0, 5.0]$ の範囲の乱数で与えた。学習には慣性項も使用した。学習反復の上限は30000回である。学習エポック数及び時間は30000回以内に学習が終了した結果の平均値である。また、ネットワークの冗長度と学習効率の関係を調べるために、49-8-8、49-12-8のネットワークに対しても本タスクを学習させた。更に、FTBP学習アルゴリズムによっても同様の学習を行った。FTBP学習においても故障集合は中間層-出力層間の単一断線故障としている。表1にはこれらの学習結果も併記してある。ここで、両学習方式において評価関数はそれぞれ E_{ME} (式(9))及び E_{FT} (式(6))を使用しているが、学習の終了条件はどちらも $\max_{f,p,k} |t_{pk} - o_{pk}^f| < 0.1$ とした。なお、(付録に示したように) Relevanceを利用した学習はその耐故障化の指針が本学習方式やFTBP学習方式とは異なることから、ここでは比較対象から除外した。

表1から耐最悪故障化学習が、FTBP学習に対し、学習に要するエポック数を大きく減少させたことが確認できる。加えて、耐最悪故障化学習はFTBP学習のように考えられ得るすべての故障状態を学習セットとして与える必要はない。そのため、1エポック内の学習量も激減する。結果として、耐最悪故障化学習は、FTBP学習と比べ、学習時間を大きく短縮した。

更に特徴的なことは、耐最悪故障化学習の効力の広

さである。FTBP学習では、ネットワークの冗長度をある程度大きくとらないと学習の成功率は上がらず、反面、冗長度を大きくするにつれて1エポック当りの学習時間が激増する。耐最悪故障化学習によれば、冗長度の小さいネットワークにおいても高い確率で耐故障性を獲得することができ、かつ、冗長度の大きなネットワークにおいてもFTBP学習に優る速度で耐故障性を獲得することが可能である。

FTBP学習では、考えられ得るすべての故障に対して結合荷重の修正が行われる。冗長度が小さい(ネットワークの自由度が少ない)場合、それらの修正が互いに競合を起こすことがあり得る。これに対し、耐最悪故障化学習の場合は、常に最悪故障状態に対する対処を行う方向に修正が進むため、効率良く耐故障性を獲得することができるのだと考えられる。

表1には、精度0.1に対する必須結合数[2]の比較結果も示した。必須結合は、FTBP学習では E (式(5)) $< \epsilon \leq E_{FT}$ (式(6))の場合に、耐最悪故障化学習では $E < \epsilon \leq E_{ME}$ (式(9))の場合に意味をもつ。よって、これらの条件を満足する学習結果を「学習の準収束」と定義し、各100回の試行のうち、学習が準収束した結果に対して必須結合数を調べ、その平均値を記している。表1から必須結合数の観点からも耐最悪故障化学習の有効性を示す結果が得られていることがわかる。

なお、表1のシミュレーションでは、経験的に学習係数及び慣性係数が選ばれている。事前に多数の試行を行い、その中で最適な学習結果を与える学習係数及び慣性係数を使用した。各学習法、各ネットワークサイズにおいて、それらの係数は異なるものとなった。また、シミュレーションはFujitsu S-7/300U model 140 (Sun Ultra 1 Model 140の互換機)にて行った。

6.2 耐故障強度の評価

6.2.1 学習対象

LSI回路においてはモジュールを多重化することにより回路を耐故障化する方法がしばしばとられる。この場合、共通モード故障に対処するためには、モジュールを単純に複製して多重化するのではなく、機能は同じであるが(負論理を用いて設計するなどして)様々な方法で構成されたモジュールを用意して複合的に多重化が行われることが多い。一般に、一つの機能を複数の表現で多重化するほどその耐故障性は向上する[1]。耐故障ニューラルネットワークにおいても、冗長ユニットが様々な内部表現を複合的に獲得するほど

その耐故障強度は高まると考えられる。

本節では、中間層が必要なが、シミュレーションにかかるコストが低く、かつ、ネットワークの内部の解析がしやすい EX-OR を例にとり評価を行う。EX-OR は線形分離が不可能な問題であり、入力層 2-中間層 2-出力層 1 のネットワークで実現されることが知られている。図 5 に EX-OR を表現するネットワークの例を示す。図 5 の場合も、中間層-出力層間の単一断線故障は非対称故障となる（図 5 の w_{11} , w_{21} の断線により、出力層ユニットの入力 $u = w_{11}h_1 + w_{21}h_2$ は必ず減少する）ので、故障をマスクするためには回路を 2 重化すればよい。よって本節では、2-4-1 のネットワークに対して耐最悪故障化学習を行い、得られる耐故障 EX-OR 回路の内部表現を調べる。本回路は 2 値出力の回路であるため、5.2 で示した最悪故障推定法が用いられる。

6.2.2 学習結果

各試行ごとに結合荷重としきい値の初期値を $[-7.0, 7.0]$ の範囲で乱数で与え、100 回の試行を繰り返した。学習係数は経験的に、結合荷重の更新に対しては 0.8, しきい値の更新に対しては 0.7 とした。学習には慣性項も使用した。慣性係数は経験的に 0.3 とした。この結果、学習反復 30000 回以内に学習が終了したのは 100 回中 37 回であった。なお評価関数は E_{ME} (式 (9)) を使用しているが、学習の終了条件は $\max_{f,p,k} |t_{pk} - o_{pk}^f| < 0.1$ を用いた。

ニューラルネットワークにおいて、各ニューロンは解空間を 2 分する超平面を構成する。EX-OR は線形分離が不可能な問題であり、これを表現するために中間層が必要であった。例えば図 6 では、中間層に $x_1 \cdot \bar{x}_2$ の領域に対して発火するニューロン A と $\bar{x}_1 \cdot x_2$ の領域に対して発火するニューロン B が現れ、出力層のニューロンがそれらを統合 (OR 操作) する

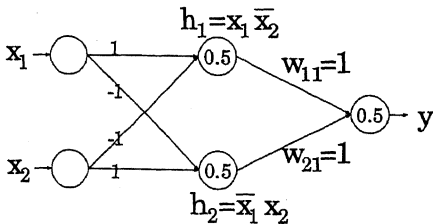


図 5 EX-OR 回路
Fig.5 EX-OR circuit.

ことにより、EX-OR ($x_1 \cdot \bar{x}_2 + \bar{x}_1 \cdot x_2 = x_1 \oplus x_2$) が実現されている。一方、図 7 に示される内部表現 (($x_1 + x_2$) · ($\bar{x}_1 + \bar{x}_2$) = $x_1 \oplus x_2$) を用いても EX-OR を実現することは可能である。ここでは、図 6 の内部表現を表現 1、図 7 の内部表現を表現 2 と呼ぶ。

本節のシミュレーションの中で、収束した 37 個のネットワークについて中間層の内部表現を調べた結果、表 2 に示された組合せにより EX-OR 回路が複合的に 2 重化されていることがわかった。これは、ニューラルネットワークでフォールトトレラント回路を作る際の大きな長所の一つである。耐最悪故障化学習によれば、複数の表現を用いて複合的にモジュールを多重化

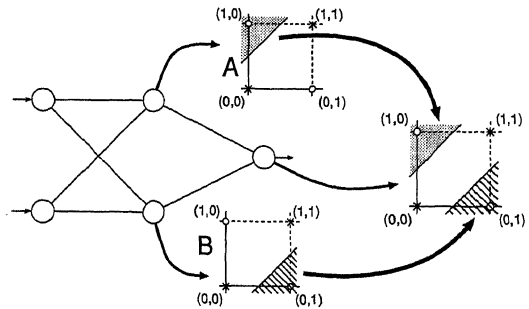


図 6 中間層の内部表現 1
Fig.6 An internal representation of hidden layer.

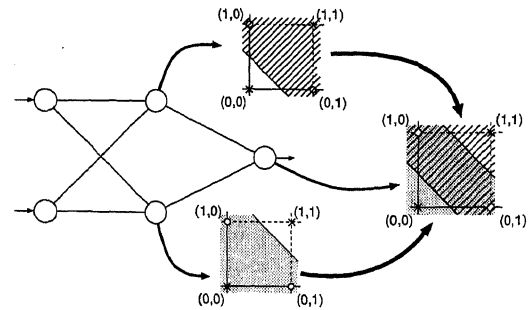


図 7 中間層の内部表現 2
Fig.7 Another internal representation of hidden layer.

表 2 内部表現の組合せ
Table 2 Internal representations of 2-4-1 networks for EX-OR.

組合せ	出現頻度
表現 1 + 表現 1	10
表現 1 + 表現 2	18
表現 2 + 表現 2	9

表3 EX-ORの学習における収束性の比較
Table 3 Comparison of the convergence properties.

	FTBP 学習	耐最悪故障化学習
ネットワークサイズ	2-4-1	2-4-1
学習成功率	12/100	37/100
平均学習エポック数	14090.6	9991.3
平均収束時間 (sec)	4.26	0.83
学習の準収束率	71/100	8/100
必須結合数 (精度 0.1)	2.18	1.88

することによって回路を耐故障化するため、モジュールを単純に複製して多重化する方法よりも耐故障強度が向上する。フォールトトレラントニューラルネットワークでは、この複合的な多重化が学習を通じて自動的に（確率的に）獲得される。

また、EX-OR 回路の学習に対しても、FTBP 学習と耐最悪故障化学習の収束性の比較を行い、表 3 にまとめた。表 3 の各項は表 1 におけるそれらに準ずる。シミュレーションの諸条件は両学習方式において同一としている。6.1 の文字認識ネットワークの学習と同様に、耐最悪故障化学習の有効性が認められる。

ただし、学習の準収束率に関しては FTBP 学習の方が大きいという結果が得られている。FTBP 学習においては、各学習エポック中で故障なしの状態に対する学習を毎回行うため、 E_{FT} (式 (6)) $< \epsilon$ を獲得することはできなかったが E (式 (5)) $< \epsilon$ については満足するという学習結果が生じやすい。一方、耐最悪故障化学習においては、 E を極小化する方向と E_{ME} (式 (9)) を極小化する方向は一般に一致しないため、学習が収束しなかった場合には $E_{ME} < \epsilon$ と $E < \epsilon$ の両者が満足されていないことが、FTBP 学習に比べて多くなるのだと推測される。なお、耐故障ネットワークを獲得する能力の高さを示すのは学習成功率の方であり、学習成功率は耐最悪故障化学習の方が大きい。

6.3 汎化能力の評価

6.3.1 学習対象

円の内外を判別するタスクを学習させる。半径 2 の円の中から 400 個の点をランダムに選び、その点が半径 1 の円の中に含まれれば 1 を出力し、その点が半径 1 の円の外であれば 0 を出力させる。ただし、400 個の点のうち、10%の点に対する教師信号を反転させることによりノイズを混入した。学習により獲得されたネットワークの入出力関係の境界がいか半径 1 の円に近づいているかを調べる。

ネットワークの入力層ユニットは二つであり、2 次

元グラフの x 座標と y 座標の値を入力する。学習時には 400 個の点の x 座標、 y 座標がアナログ値で入力される。出力層は円の内外を判別するための 1 ユニットののみとなる。

通常の誤差逆伝搬学習では、中間層のユニット数がネットワークの汎化能力に大きな影響を与える。中間層のユニット数が少なすぎると正しい入出力関係を習得することはできない。逆に中間層のユニット数が多すぎると、ノイズを含んだ入出力関係を習得（オーバーフィッティング）してしまう [5]。ネットワークの汎化能力を引き出すためには中間層のユニット数を適切に選ばなければならないが、最適な中間層ユニット数を事前に求めることは一般に難しいとされている。

本節では円の内外を判別するタスクを耐最悪故障化学習により学習させ、中間層のユニット数とネットワークの汎化能力との関係を調べる。本タスクは 2 値出力となるので 5.2 で示した最悪故障推定法が用いられる。

6.3.2 学習結果

はじめに、中間層のユニット数を変えて、通常の誤差逆伝搬学習による学習を行った。結合荷重としきい値の初期値は $[-5.0, 5.0]$ の範囲の乱数で与えた。学習係数は経験的に、結合荷重の更新に対しては 0.3、しきい値の更新に対しては 0.2 とした。学習には慣性項も使用した。慣性係数は経験的に 0.05 とした。本タスクの学習セットはノイズを含んでいるので学習を繰り返しても収束することはない。100000 回の学習を終えた時点で学習を完了したこととし、その時点のネットワークが獲得している入出力関係を調べた。中間層ユニット数が 10 個のときの結果を図 8 に示す。黒く塗られている部分が出力が 0.5 以上となる領域である。図には学習セットとした 400 個の点も示してある（○印の点に対する教師信号が 1、×印の点に対応する教師信号が 0 である）。図 8 から、ネットワークはノイズを含んだ入出力関係を学習してしまっていることがわかる。

一方、同じく 2-10-1 のネットワークに対し、同じ学習セットを用いて、耐最悪故障化学習により同タスクを学習させた結果を図 9 に示す。学習の諸条件も誤差逆伝搬学習と同じである。耐最悪故障化学習によれば、入出力関係の境界は半径 1 の円に近く、その汎化能力の高さが見てとれる。

前節までのシミュレーションで、耐最悪故障化学習が回路を多重化することにより耐故障化を達成してい

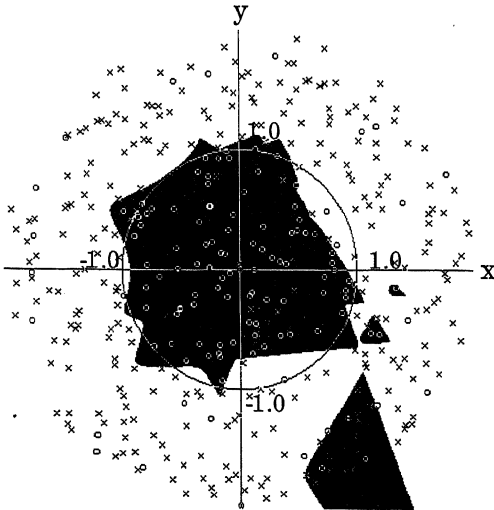


図8 半径1の円を学習したネットワークの出力分布（誤差逆伝搬学習，中間層10）
 Fig.8 Output distribution of a network that learned circle with radius 1 (back propagation algorithm, 10 hidden units).

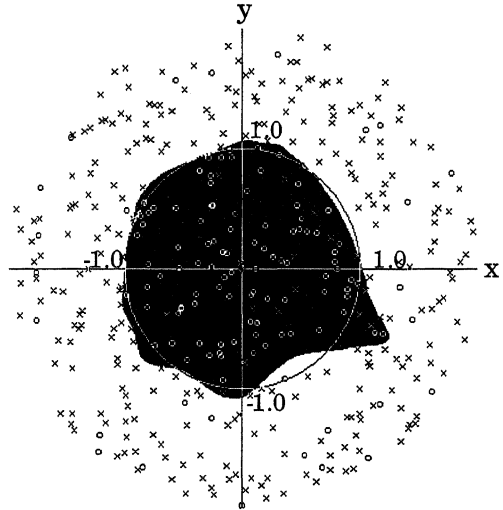


図10 半径1の円を学習したネットワークの出力分布（耐最悪故障化学習，中間層40）
 Fig.10 Output distribution of a network that learned circle with radius 1 (the proposed algorithm, 40 hidden units).

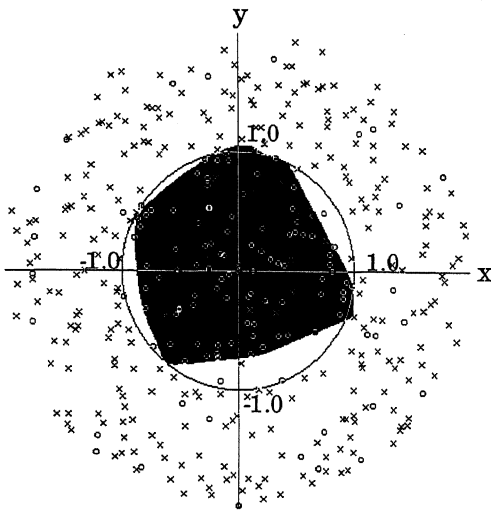


図9 半径1の円を学習したネットワークの出力分布（耐最悪故障化学習，中間層10）
 Fig.9 Output distribution of a network that learned circle with radius 1 (the proposed algorithm, 10 hidden units).

ち、耐最悪故障化学習においては、中間層のユニット数は、通常の誤差逆伝搬学習を用いるときに比べ、見かけ上約1/3になっていると考えられる。これを考慮して、2-40-1のネットワークに対しても、同じ学習セット、学習条件を用いて、耐最悪故障化学習により同タスクを学習させてみた。結果が図10である。中間層ユニット数が10個の場合と同程度の汎化能力が得られていることがわかる。

以上から、耐最悪故障化学習が FTBP 学習と同じく、ネットワークの汎化能力をも向上させることが確認された。更に、耐最悪故障化学習を行うことによる汎化能力の向上は、回路の多重化による効果を差し引いてもなお余りあるものであった。

通常の誤差逆伝搬学習の場合、中間層内の各々のニューロンが勝手に任意の内部表現をとることが可能なため、ノイズに過敏に反応してしまい、オーバーフィッティングが起こる。耐最悪故障化学習によれば、故障時に備え、複数の中間層ユニットがお互いに相補うように内部表現を獲得する。この結果、ある中間層ユニットがノイズに過敏に反応しすぎてしまうことを他の中間層ユニットが抑制するのだと推測される。

なお、図8～図10は、それぞれネットワークの結合荷重としきい値の初期値を変えて100回の試行を行

ることが確かめられた。本論文では中間層-出力層間の単一断線故障を仮定しているので、回路は3重化により耐故障化されているものと推測される [1]。すなわ

い、出力分布の傾向を調べた上で、最も高頻度で得られた結果の中から無作為に選んだものである。

7. む す び

ニューラルネットワークを利用して効率的にフォールトトレラント回路を実現する方法を示した。通常の誤差逆伝搬アルゴリズムを用いて各学習セットに対する最悪故障時の出力を教師信号に一致させることにより耐故障性を獲得する耐最悪故障化学習アルゴリズムを提案した。更に、最悪故障出力を高速に計算するための最悪故障推定法を示した。

実際に7×7画素の八つのアルファベットに対して文字認識を行うタスクをシミュレートし、本学習方式により耐故障性をもつ回路を効率良く獲得できることを確認した。EX-OR 回路を獲得するシミュレーションでは、本学習方式が、様々な内部表現を用いて回路を複合的に多重化(2重化)することにより耐故障性を達成していることが確認された。円の内と外を判別するタスクを学習させるシミュレーションからは、本学習方式が優れた汎化能力をも呈するという結果が得られた。

現段階では中間層-出力層のみの断線に対する耐故障性を獲得するアルゴリズムを提供した。本アルゴリズムは多層ニューラルネットワークのすべての結合の断線に対する方式に拡張が可能である。これについてはアルゴリズム等を体系的にまとめ、近日中に別途報告する予定である。

更に、大規模回路の耐故障化に対する本学習方式の有効性についても調査していく。

文 献

- [1] 南谷 崇, フォールトトレラントコンピュータ, オーム社, 1991.
- [2] 丹 康雄, 南谷 崇, “フォールトトレランスを有する階層型ニューラルネットワークとその性質,” 信学論 (D-I), vol.J76-D-I, no.7, pp.380-389, July 1993.
- [3] N.C. Hammadi and H. Ito, “A learning algorithm for fault tolerant feedforward neural networks,” IEICE Trans. Inf.& Syst., vol.E80-D, no.1, pp.21-27, Jan. 1997.
- [4] 都筑輝泰, 西垣正勝, 曾我正和, “最悪故障推定法による耐故障ニューラルネットワークの自己組織化,” 第11回回路とシステム(軽井沢) ワークショップ論文集, pp.77-82, April 1998.
- [5] 上坂吉則, “ニューラルネットと学習可能性,” 信学誌, vol.74, no.9, pp.943-948, 1991.

付 録

[結合荷重の制限と耐故障化の関係]

ニューラルネットワークで表 A.1 の入出力関係を学習する場合を考える。

この場合、単一断線故障に対する耐故障性を与えるためには、ネットワークの入力を2重化($x = x_1 = x_2$)し、図 A.1 のように構成すればよい。したがって、耐故障化学習は図 A.1 のネットワークに表 A.2 の入出力関係を覚えさせることと等価になる。

図 A.1 のネットワークを FTBP 学習により耐故障化する場合、その誤差関数は

$$E_{FT} = \sum_f \sum_p (y - o_p^f)^2 \quad (\text{A.1})$$

である。ここで、 $o_p = \text{sigm}(w_1 x_1 + w_2 x_2 + \theta)$ は第 p パターンを入力した際のネットワークの出力であり、 f は任意の単一断線故障、 o_p^f は f が起こったときのネットワークの出力である。したがって、学習方程式は各パターン p 、各故障 f に対して

$$\begin{aligned} w_i^{(m+1)} &= w_i^{(m)} - \alpha \frac{\partial E}{\partial w_i^f} \\ &= w_i^{(m)} + 2\alpha(y - o_p^f) o_p^f (1 - o_p^f) x_i \end{aligned} \quad (\text{A.2})$$

表 A.1 入出力関係
Table A.1 Truth table.

x	y
0	0
1	1

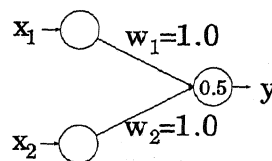


図 A.1 2重化したニューラルネットワーク
Fig. A.1 Duplicated neural network.

表 A.2 2重化された入出力関係
Table A.2 Truth table for duplicated network.

pattern p	input		output y
	x_1	x_2	
1	0	0	0
2	1	1	1

と導かれる。ここで、 w_i^f は結合 w_i に故障 f が発生したことを表す。また、 $\alpha > 0$ は学習係数、 m は学習反復回数である。

式 (A.2) において、 $0 < \sigma_f < 1$ 、 $x_i = 1$ または 0 、 $y = 1$ または 0 であることから、 $y = 1$ 、 $x_1 = x_2 = 1$ のときには $w_i^{(m+1)} > w_i^{(m)}$ であることが示される。すなわち、耐故障性を得るためには、結合荷重を大きくする方向に学習を進めなければいけないことが十分にあり得る。これに対し、Hammadi らの耐故障化アルゴリズムは（影響度 R の大きい結合枝に対する）結合荷重を小さく制限することのみにより耐故障性を得ようとするものであり、耐故障化の指針が FTBP 学習のものとは異なっているといえる。

（平成 11 年 2 月 10 日受付，5 月 21 日再受付）



西垣 正勝（正員）

平 2 静岡大・工・光電機械卒。平 4 同大大学院修士課程了。平 7 同大大学院博士課程了。日本学術振興会特別研究員 (PD) を経て、平 8 静岡大情報学部助手、平 11 同講師、現在に至る。博士 (工学)。回路シミュレーション、ニューラルネットワーク、

通信セキュリティなどに関する研究に従事。IEEE 会員。



都筑 輝泰

平 9 静岡大・工・情報知識卒。平 11 同大大学院博士課程前期理工学研究科計算機工学専攻了。同年シャープ (株) 入社。現在、赤外線通信関連の LSI 開発に従事。在学中、ニューラルネットワークとフォールトトレランスに関する研究に従事。



曽我 正和（正員）

昭 33 京大・工・電子卒。昭 35 同大大学院修士課程電子了。昭 35～平 8 三菱電機計算機製作所、情報電子研究所、本社開発本部。平 8 静岡大学情報学部教授、平 11 岩手県立大学ソフトウェア情報学部教授、現在に至る。博士 (工学)。汎用計算機、制御

用計算機、制御用システムの開発に従事。フォールトトレラントシステム、セキュリティシステムに関する研究に従事。