

内部表現の分散化によるニューラルネットワークの耐故障化の一手法

1 D-3

名古 和行 丹 康雄 南谷 崇

東京工業大学 工学部

1 はじめに

階層型ニューラルネットワーク [1, 2] における故障を考えたとき、ネットワークの構造に冗長性があり、情報を分散して持っていれば耐故障性は向上すると考えられる。しかしながら、単に構造を冗長化したネットワークを通常の誤差逆伝搬 (Error Back Propagation) アルゴリズム [1] (以下 BP) 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。BP は誤差評価関数を極小にするようにパラメータを修正していくだけであるため、情報の分散の保証は何もない。

本稿では、一度 BP によって学習させたネットワークに対し、特定の重みの初期化と再学習の繰り返しによりネットワークの内部表現を分散させ、耐故障性を向上させる手法を提案する。また、実験によってその有効性を示す。

2 ネットワークモデル

本稿で扱うニューラルネットワークは入力層、隠れ層、出力層の3層の階層型ネットワークとし、層を飛び越えた結合は認めず、信号の伝搬は入力から出力への一方である (フィードバック結合がない) とする。各ユニット (素子) の出力は、入力の荷重和に対しシグモイド (S 字状) 関数 (式 (2)) を用いることにより決定される。このようなユニットを層状に配置し、層の間にユニットどうしの結合を設けたネットワークが階層型ネットワークである。

$$u = \sum_{i=1}^n w_i x_i - \theta \tag{1}$$

$$f(x) = \frac{1}{1 + \exp(-u)} \tag{2}$$

但し、 x はユニットへの入力、 w はその重み、 θ は閾値、 u は入力の総和、 f は出力関数を表す。

3 耐故障性の評価基準

本稿では故障を単一の結合の断線 (重みの 0 縮退) のみに限定する。

ある結合を切断した結果、ネットワークが誤りを起こすとき、その結合を必須結合とよぶことにする。この数をネットワークの耐故障性の評価基準にし、必須結合が少ないほど耐故障性は高いと考える。誤りの基準は問題に依存するため、ここでは特に規定しない。

4 分散表現と局所表現

分散表現と局所表現を次のように定義する。

定義 1 ネットワーク内部の分散表現

どの入力パターンに対しても多数の隠れユニットからの入力によりネットワークの出力が決定され、それぞれの役割が均等であること。

定義 2 ネットワーク内部の局所表現

ある入力パターンに対し、ネットワークの出力が少数の隠れユニットからの入力によって決定されること。

A method to make fault tolerant neural networks by distributing internal representation

Kazuyuki Nako, Yasuo Tan and Takashi Nanya

Faculty of Engineering, Tokyo Institute of Technology

隠れユニットとは隠れ層にあるユニットのことである。

ネットワークが局所表現されている場合、特定のユニットからの情報の欠如や誤りにより、ある入力に対し誤りを起こす。しかし、分散表現では同じ情報を多数のユニットが持っているため、誤りがマスクされる可能性がある。

5 BP とネットワークの局所性

階層型ネットワークの学習法として BP が有名である。これは、入力パターンを示したときの出力の誤差をネットワークの逆方向に伝搬させ、誤差評価関数を極小にするように重みと閾値を修正する学習方法である。適当な局所解に収束すれば良いため、情報が分散する保証はない。一般的に、隠れ層に冗長性を持たせても学習は均等に進まず、特定の隠れユニットの出力がネットワークの出力を決定するようになる (図 1)。

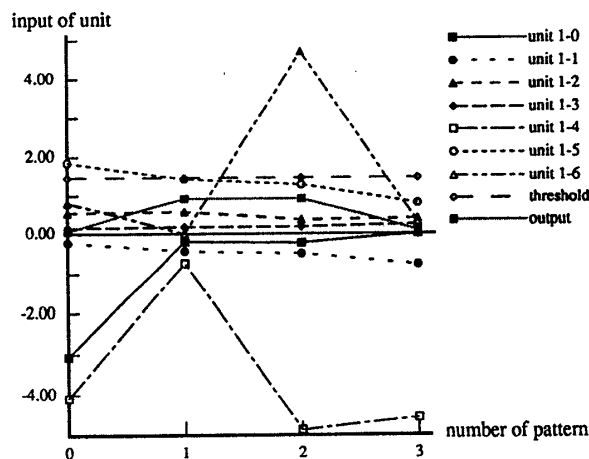


図 1: 出力ユニットの入力の分布 1 [2-7-1 xor]

図 1 は、各層のユニットの数が (入力から) 2-7-1 のネットワークにおいて、通常の BP で xor 問題を学習させたときの出力ユニットの入力の分布を表している。凡例の上から 7 つが隠れ層からの入力である。それぞれの隠れユニットからの入力は重み付けをしてある。グラフの閾値は $-\theta$ を表している。学習率は 0.1、慣性率は 0.9 である。入力パターンの提示は順に行い、重みは逐次更新する。学習の終了条件は、各出力ユニットと教師パターンとの誤差の 2 乗の最大値が 0.01 以下になることとした。

入力の大きさが不均等であり、ネットワークの出力に対して支配的な入力と、ほとんど影響がない入力が存在する。つまり、ネットワークが局所表現されていると同時に、無駄な部分があるといえる。

6 新しい手法

本稿で提案するアルゴリズムを図 2 に示す。まず学習を行い、必須結合の重みを初期化し、再び学習を行う。以後これを繰り返す。必須結合であるということは、その結合を通る信号がネットワークに対して支配的であり、重要な情報を持つことを意味する。故に、この重みを初期化することにより、他の重みに情報が分散することが期待できる。

重みを初期化するのは出力ユニットの入力の重みだけとする。隠

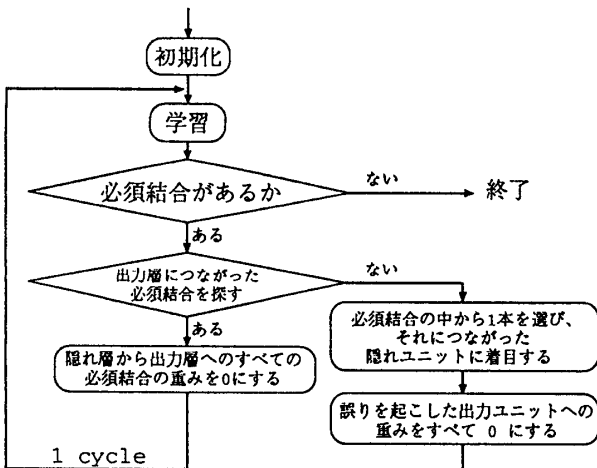


図 2: 本稿で提案するアルゴリズム

れユニットの入力の重みまで初期化すると、図 1 に示された隠れユニットの出力の形状まで学習することになる。前節で説明した無駄な部分を有効に活用し、ネットワーク全体の役割が同等になるようにするためには、出力ユニットの入力の大きさが均等になるように学習を進めれば良いので、形を変えることは意味がない。このため、出力層につながった必須結合が存在しない場合、入力層から隠れ層への必須結合の重みは初期化せず、その必須結合につながった隠れユニットから出力層への重みを初期化する。

ネットワーク出力に対して支配的である必須結合の重みは初期化されるため、図 1 にあるような大きな入力存在できない。このため、初期化された重みが成長する間に他の値の小さな重みが成長し、入力が均等になる方向に学習が進むと考えられる。またこのとき、大きさが小さく、無駄になっていた重みは再学習により大きくなり、ネットワークの出力に貢献するようになる。

7 実験

図 3 は図 1 で用いたネットワークに本アルゴリズムを適用したものである。学習方法、学習パラメータ、重みの初期値は同一である。

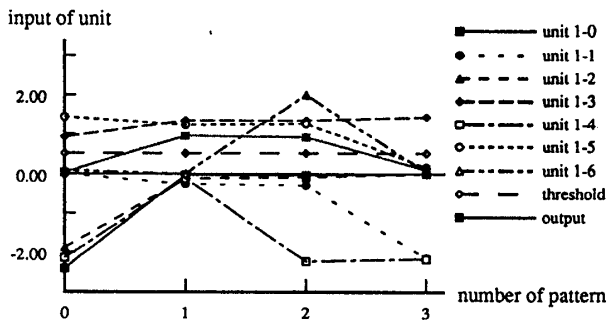


図 3: 出力ユニットの入力の分布 2 [2-7-1 xor]

閾値は常に 1 を出力するユニットと重みで表現し、故障の対象にした。今回は 2 値問題のみを扱うため、故障時の誤り判定では 0.5 未満を 0、0.5 以上を 1 と判断する。

図 3 は図 1 と比較して、入力の大きさの偏りがなく、一つのパターンが複数のユニットで表現されており、内部表現の分散化が進んだことが分かる。

図 4 は本アルゴリズムにおける必須結合数の推移である。再学習を繰り返すにつれて耐故障性が向上し、最終的に任意の単一断線故障をマスクするようになった。

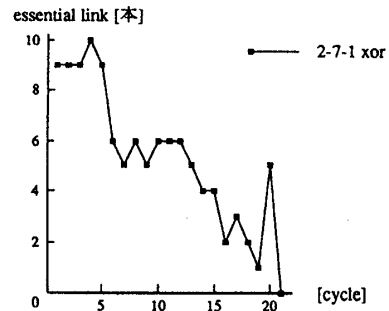


図 4: 本アルゴリズムにおける必須結合数の推移 [2-7-1 xor]

同様に 4bit parity、入力 5bit 出力 5bit の random mapping 問題についても、適当に隠れ層に冗長性を持たせれば (前者は 4-12-1、後者は 5-10-5 の構成) 任意の単一断線故障をマスクすることが確認できた。また、実験を行った 3 つの問題のすべてにおいて、隠れユニットの数を増加させるにつれ、より少ない再学習回数で故障をマスクするようになった。

8 考察

本実験では隠れユニットの数を、いずれも対象とする問題を学習可能な最小のネットワークの 3 倍程度以上用意すれば、任意の単一断線故障をマスクすることができた。しかしながら、これは 2 値問題を扱い、故障時の許容誤差を大きくとったときの結果であり、許容誤差を小さくした場合、故障による誤差を小さくしなければならないため、さらに多くの隠れユニットが必要になる。

耐故障性を高めるために隠れユニットを増やすとき、単一故障だけを考えるならば問題ないが、複数の故障を考えた場合、回路が大きくなれば、それだけ故障が発生する確率も高くなるため、実際に耐故障性が向上するかはわからない。故障する確率も考慮に入れた耐故障性の指標を考えなければならない。

本アルゴリズムにより得られたネットワークは隠れユニットの固定故障にも強いことが期待できる。しかし、完全にマスクするためには、隠れユニットの出力の重みのうち、固定故障を起こすことにより誤りを生じた出力ユニットへの重みを初期化するなど、アルゴリズムに何らかの工夫が必要である。0 固定故障に限って言えば、そのユニットから出力層へのすべての結合の断線と等価にみることが出来る。一般に、隠れユニットから出力層への結合はネットワークの出力に対し互いに独立に作用するため、それぞれの断線は一つの出力ユニットにしか悪影響を及ぼさない。そのため、断線故障がマスクされれば 0 固定故障もマスクされる。

9 まとめ

重要な情報を持った重みを初期化し、再学習を行うことにより情報を分散させ、耐故障性を向上させる方法を提案した。実験では 2 値問題を扱い、単一の断線故障をマスクできることを示した。

本稿では単一断線故障のみを想定していたが、固定故障や複数の故障が発生した場合の耐故障性についても今後検討する必要がある。

参考文献

[1] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation", Parallel Distributed Processing, Vol. 1, eds. J. L. McClelland, D. E. Rumelhart, and The PDP Research group. Cambridge, MA : MIT Press, pp.318-362, (1986).
 [2] 麻生 英樹, "ニューラルネットワーク情報処理", 産業図書, (1988).