

しきい値ゆらぎを利用した多層パーセプトロンの並列相関学習法とその特性

5H-2

西田 啓蔵 三谷 光照 大堀 隆文 渡辺 一央
北海道工業大学 電気工学科

1 まえがき

誤差逆伝搬 (BP) 法とは異なる, 小振幅, 不偏かつ互いに無相関なしきい値ゆらぎを直接学習に利用する並列相関学習 (PCL) 法を提案し, その理論的検討を行う。また, XOR課題に対するシミュレーションにより学習動作の確認をし, BP法との計算時間の比較を行ったので報告する。

2 ニューロンモデル

各ニューロンのしきい値に「ゆらぎ」が存在する, 情報伝達のための諸量が入力標本 p のみならず時刻 t に関して変化・変動している, ことを除いては, 従来のニューロンモデルと基本的には同じである。ニューロン k の入力標本 p に対する入出力関係は, 以下のように記述できる。ただし, $\langle \cdot \rangle$: 期待値 (時間平均), $f\{\cdot\}$: 入出力関数, T : 転置を表し, 定義されたベクトルは列ベクトルとする。

- ニューロン k への結合係数ベクトル : w_k
- シナプス前ニューロン群出力ベクトル : $q_k(p, t)$
- 入力総和 : $u_k(p, t) = w_k^T q_k(p, t)$
- 入力総和の平均 : $a_k(p) = \langle u_k(p, t) \rangle$
- 入力総和のゆらぎ成分 : $m_k(t) = u_k(p, t) - a_k(p)$
- しきい値ゆらぎ成分 : $n_k(t)$
- ニューロン出力 : $r_k(p, t) = f\{u_k(p, t) + n_k(t)\}$

3 出力誤差の近似表現

情報伝達が一方方向の feedforward 形のネットワークを考える。ネットワークの出力ベクトルに関して, 微分可能な出力誤差関数 $e(p, t)$ を考える。微分可能であれば, その関数形にはこだわらない。各ユニットの $a_k(p), m_k(t), n_k(t)$ からなるベクトルを, $a(p), m(t), n(t)$ によって表す。ここで,

$$\text{条件 1: } n(t) \approx 0 \tag{1}$$

を設定すると, $m_k(t)$ はユニット k へ情報を伝達しているユニット h 群のしきい値ゆらぎ $n_h(t)$ 群の線形和で近似できるので, $m(t)$ は係数行列 C を用いて次式のように表せる。

$$m(t) \approx Cn(t) \tag{2}$$

式(2)により, $m(t)$ は, $n(t)$ の1次形式で近似できる従属変数で, $n(t) = 0$ の時, $m(t) = 0$ となるから, $e(p, t)$ は, 独立変数 $n(t)$ のみに関するテーラー展開を用いて次式で表される。

$$e(p, t) = e\{a(p) + m(t) + n(t)\} \approx e\{a(p)\} + \frac{\partial e\{a(p)\}}{\partial a(p)} n(t) \tag{3}$$

4 並列相関学習法

4.1 最小化原理

全ユニットのゆらぎが存在しない状態 $n(t) = 0$ の出力誤差関数値 $e(p) = e\{a(p)\}$ を, ユニット k への結合係数 w_k に関する最急降下法によって最小化する。ここで, 結合係数の修正 $\Delta_p w_k$ は, 次式で与えられる。

$$\Delta_p w_k^T = -\epsilon \frac{\partial e(p)}{\partial w_k} = -\epsilon \frac{\partial e(p)}{\partial a_k(p)} \cdot \frac{\partial a_k(p)}{\partial w_k} \tag{4}$$

A Parallel Correlation Learning by Threshold Fluctuation for Multi-layered Perceptrons and The Characteristic
Keizo NISHIDA, Mitsuaki MITANI, Takahumi OOHORI, Kazuhisa WATANABE
Hokkaido Institute of Technology, Teine-ku, Sapporo, JAPAN

4.2 偏微分量の推定

式(3)の両辺に, 右辺から $n(t)^T$ を乗じ, その期待値を求めると, 次式を得る。

$$\langle e(p, t)n(t)^T \rangle \approx e(p) \langle n(t)^T \rangle + \frac{\partial e(p)}{\partial a(p)} \langle n(t)n(t)^T \rangle \tag{5}$$

ここで, 以下の条件を設定する。

$$\text{条件 2: } \langle n(t) \rangle = 0 \tag{6}$$

条件 3: 相互相関がない。

$$\text{すなわち, } \langle n(t)n(t)^T \rangle = \text{diag}\{\langle n_k(t)^2 \rangle\} \tag{7}$$

ただし $\text{diag}\{\cdot\}$ は, 対角行列を表す。式(5)に式(6)~(7)を代入することにより,

$$\frac{\partial e(p)}{\partial a(p)} \approx \langle e(p, t)n(t)^T \rangle \text{diag}\{\langle n_k(t)^2 \rangle^{-1}\} \tag{8}$$

$$\frac{\partial e(p)}{\partial a_k(p)} \approx \frac{\langle e(p, t)n_k(t) \rangle}{\langle n_k(t)^2 \rangle} \tag{8}$$

となる。さらに, 式(4)中の偏微分量 $\partial a_k(p)/\partial w_k$ は, 以下のように書き表せる。

$$\frac{\partial a_k(p)}{\partial w_k} \approx \frac{\partial \langle w_k^T q_k(p, t) \rangle}{\partial w_k} = \langle q_k(p, t)^T \rangle \tag{9}$$

4.3 学習則 (結合係数修正式)

式(8),(9)を式(4)に代入することによって, 微分処理や逆伝搬処理を含まない簡潔な学習則として, 次式を得る。

$$\Delta_p w_k^T = -\epsilon \frac{\langle e(p, t)n_k(t) \rangle}{\langle n_k(t)^2 \rangle} \langle q_k(p, t)^T \rangle \tag{10}$$

以上の理論的検討から, 本学習法は,

1. 微分処理や逆伝搬処理をすることなくすべての結合係数を並列かつ同時に修正可能,
2. ユニットの入出力非線形変換の関数形に関わる学習式の修正なしに学習可能,
3. 層構造を意識することなく学習処理可能, などの特徴を有する。

5 シミュレーション

小振幅, 不偏かつ互いに無相関なしきい値ゆらぎ群を用いた XOR課題に対するシミュレーションから, 本学習法の学習可能性を確認できた (Fig.1)。また, 本学習法の想起も含めた全体の演算量は, ネットワーク規模が大きくなるに従い, BP法に比べておよそ $T_p/2$ (大規模) ~ T_p (小規模) 倍となるが, 結合係数修正演算量は BP法とほぼ同等となった (Fig.2)。

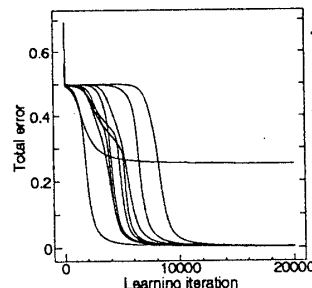


Fig.1 学習特性

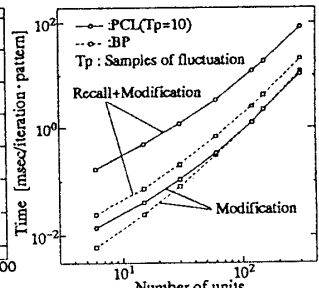


Fig.2 PCL法とBP法の計算時間の比較

6 むすび

本報告では, 従来の BP法と異なる, 汎用的な並列相関学習法を理論的に提案し, XOR課題に対するシミュレーションによってその学習動作を確認した。また, 本学習法は, BP法と同等の学習性能を有することがわかった。今後, 本学習法の性能支配要因, 適用可能範囲などについて検討していく予定である。