1997年度日本オペレーションズ・リサーチ学会
春季研究発表会

階層型ニューラルネットワークに対する逐次射影法の改良(その2)

京都大学工学研究科　*巽 啓司　TATSUMI Keiji
京都大学工学研究科　福島 雅夫　FUKUSHIMA Masao

1. はじめに

著者らは階層型ニューラルネットワークに対するon-line型の学習アルゴリズムである**逐次射影法**を提案してきた[2,3,4,5]. とくに[5]では, 使用する階層型ニューラルネットワークに若干の修正を加え, 非線形性が強いため扱いが困難であった部分問題をより扱いやすい形に再定式化することで逐次射影法を拡張した. さらに計算機実験でこの方法が従来のBP法[1]に比べ反復回数を大幅に軽減できることやアルゴリズムの安定性を確認したが, 厳密に部分問題を解いているため, 大規模な問題に対してはかなりの計算時間を必要とすることも分かった. そこで本報告ではこの問題を解消するため, 問題の構造を有効に利用した反復法を用いて部分問題を近似的に解くことにより, 逐次射影法を改良することを提案する.

2. 階層型ニューラルネットワーク

階層型ニューラルネットワークとは多入力1出力のユニットを階層状に構成したものである. 以下, 本報告では, 3層型ニューラルネットワークを考え, 各ユニットの出力関数 $f(x)$ としてはシグモイド関数 $s(x) = (1-exp(-\lambda x))/(1+exp(-\lambda x))$ を用いる. 各層のユニット数は入力層 L, 中間層 M, 出力層 N とする. ニューラルネットッワークへの入力を $\boldsymbol{x} \in R^L$ と表し, 中間層からの出力を $\boldsymbol{y} \in R^M$, 出力層からの出力を $\boldsymbol{z} \in R^N$ と表す. 入力層, 中間層間の結合荷重を $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_M\}$ で, 中間層, 出力層間の結合荷重を $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_N\}$ で表すことにする. ただし, $\boldsymbol{u}_i$ は入力層と中間層ユニット i 間の結合荷重を表すベクトル $\boldsymbol{u}_i = (u_{1i}, u_{2i}, \cdots, u_{Li})^\top$ であり, $\boldsymbol{v}_j$ は中間層と出力層ユニット j 間の結合荷重を表すベクトル $\boldsymbol{v}_j = (v_{1j}, v_{2j}, \cdots, u_{Mj})^\top$ である. また, 各ユニットの閾値は常に -1 の入力にかかる重みと考え, 結合荷重と同一視することにし, 全結合荷重はベクトル $\boldsymbol{w} = (\boldsymbol{u}_1^\top, \boldsymbol{u}_2^\top, \cdots, \boldsymbol{u}_M^\top, \boldsymbol{v}_1^\top, \cdots, \boldsymbol{v}_N^\top)^\top$ で表す. さらに, ネットワークの出力つまり出力層の各ユニット i からの出力を

$$z_i(\boldsymbol{w}, \boldsymbol{x}^p) := f({\boldsymbol{v}_i}^\top \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{x}^p))$$

で表す. ただし,

$$\boldsymbol{f}(\boldsymbol{u}, \boldsymbol{x}^p) := (f({\boldsymbol{u}_1}^\top \boldsymbol{x}^p), \cdots, f({\boldsymbol{u}_M}^\top \boldsymbol{x}^p))^\top$$

である.

3. 逐次射影法

階層型ニューラルネットワークの学習とは与えられた P 個の入出力の組 $\{\boldsymbol{x}^p, \boldsymbol{t}^p\}$ を実現するような結合荷重 $\boldsymbol{w}$ を求めることである. ここで, $\boldsymbol{x}^p, \boldsymbol{t}^p$ はそれぞれ入力パターン, 教師信号と呼ばれる. この学習は実質的にパターン分離問題であり, 教師信号 $\boldsymbol{t}^p$ の要素は $1, -1$ のように2値が用いられる. このとき, 学習の目的は $1, -1$ そのものを出力することではなく, $-1 \le T_l \le T_u \le 1$ なる定数 T_l, T_u を用いて, 各ユニットからの出力を T_l 以下, T_u 以上とすることである. したがって, 用いるネットワークの結合荷重 $\boldsymbol{w}$ が, 教師信号の数 P に対し十分に大きな次元であるとすれば, 上記の入出力を満たす $\boldsymbol{w}$ は常に存在すると考えられるから, ニューラルネットワークの学習は以下の非線形連立不等式を $\boldsymbol{w}$ について解くことと等価になる.

$$z_i(\boldsymbol{w}, \boldsymbol{x}^p) \le t_i^p, \quad \begin{array}{l} p \in \{1, \cdots, P\}, \\ i \in \{1, \cdots, N\}. \end{array} \tag{1}$$

ただし, 便宜上(1)式は不等号の向きを揃えてある. 以下では, この連立不等式を解くことを考える.

ここで(1)式をさらに解きやすくするため以下の変形を行う. 出力関数 f は狭義単調増加関数であるから, (1)式の連立不等式は次のように書きかえることができる.

$${\boldsymbol{v}_i}^\top \boldsymbol{f}(\boldsymbol{u}, \boldsymbol{x}^p) \le f^{-1}(t_i^p), \quad \begin{array}{l} p \in \{1, \cdots, P\}, \\ i \in \{1, \cdots, N\}. \end{array} \tag{2}$$

また, さらに中間層のユニットの出力関数として, $f(x)$ を少し修正した関数

$$h(x) := \begin{cases} f(X_u) + f'(X_u)(x - X_u), & (x \ge X_u) \\ f(x), & (X_l < x < X_u) \\ f(X_l) + f'(X_l)(x - X_l), & (x \le X_l) \end{cases}$$

を用い, (2)式の代わりに

$$z_i^p(\boldsymbol{w}) := {\boldsymbol{v}_i}^\top \boldsymbol{h}(\boldsymbol{u}, \boldsymbol{x}^p) \le f^{-1}(t_i^p), \quad \begin{array}{l} p \in \{1, \cdots, P\}, \\ i \in \{1, \cdots, N\}. \end{array} \tag{3}$$

を解くことにする. ただし,

$$\boldsymbol{h}(\boldsymbol{u}, \boldsymbol{x}^p) := (h({\boldsymbol{u}_1}^\top \boldsymbol{x}^p), \cdots, h({\boldsymbol{u}_M}^\top \boldsymbol{x}^p))^\top$$

である．ここで，パターン p に関する (3) 式を $\boldsymbol{w}^l$ の現在の値において線形近似した制約条件をもつ２次計画部分問題を次式で定義する．

$$\begin{aligned} \min \quad & \frac{1}{2}\|\boldsymbol{w}-\boldsymbol{w}^l\|^2+\frac{1}{2}Cr^2 \\ \text{s.t.} \quad & \xi_i^p(\boldsymbol{w}^l)+\nabla\xi_i^p(\boldsymbol{w}^l)^\top(\boldsymbol{w}-\boldsymbol{w}^l)+r \le f^{-1}(t_i^p) \\ & \qquad\qquad\qquad\qquad i\in\{1,\cdots,N\}, \\ & h^{-1}(-1)\le \boldsymbol{u}_j{}^\top \boldsymbol{x}^p \le h^{-1}(1),\ \ j\in\{1,\cdots,M\}. \end{aligned} \tag{4}$$

なお，変数は $\boldsymbol{w}$ と r であり，C は正の定数である．また，関数 $h(x)$ の出力値を $-1\le h(x)\le 1$ に抑える制約条件を加えてある．

本報告で提案する逐次射影法 (Modified Successive Projection Method, MSPM) は，この各パターンごとの部分問題を逐次的に解くことにより結合荷重の列 $\{\boldsymbol{w}^l\}$ を生成する方法である．

MSPM

Step 1 結合荷重の初期値 $\boldsymbol{w}^0$ を選び，$l:=0, p:=1$ とする．

Step 2 パターン p に関する部分問題 (4) の解 $\boldsymbol{w}^*$ を求め，$\boldsymbol{w}^{l+1}:=\alpha(\boldsymbol{w}^*-\boldsymbol{w}^l)+\boldsymbol{w}^l$ とする．ここで，$\alpha\in(0,2)$ は緩和パラメータである．

Step 3 終了条件が満たされなければ，$p:=p+1 (\mathrm{mod}\ P)$, $l:=l+1$ として **Step 2** に戻る．

この方法は，on-line 型の学習アルゴリズムであり，現在の $\boldsymbol{w}^l$ から各パターンに対する部分問題 (4) の制約条件が表す半空間への距離に基づいて結合荷重の更新を行う方法である．それゆえ，学習の進行と共に適応的に結合荷重の更新が行われることが期待できる．

4. 部分問題の解法

部分問題 (4) を解くため，問題の特殊構造を考慮して，区間制約を持つ２次計画問題に対して開発された SOR 法 (逐次過緩和法)[6] を用いた．一般に SOR 法は，変数の次元に比べ制約条件が少ない場合に有効な方法である．上で述べた３層型ニューラルネットワークの場合，変数 $\boldsymbol{w}$ の次元は $NM+ML$ であり，部分問題の制約の数は $N+M$ であるから，その特徴を備えている．また，SOR 法の収束判定条件を調整し，部分問題を近似的に解くことにより，MSPM 全体の計算時間の軽減をはかった．

5. 実験結果

MSPM, on-line BP と BP モーメント法をいくつかの問題に適用した．ここでは，排他的論理和 (XOR) を実現する問題 A と，ランダムに生成した３入力３出力の８個の入出力の組み合わせを学習する問題 B に適用したときの結果を示す (表 1)．入力パターンの順番については，毎回固定した順番を用いる cyclic な方法と，各反復ごとにパターンを入力する順序をランダムに定める almost cyclic な方法の二つを用いた．どちらの方法においても，全パターンが一通り入力されたとき一回の反復が行われたものとした．ここで，"#Ite.","CPUsec" はそれぞれ，50 個の異なる初期値に対する試行の平均反復回数，平均学習時間を表し，また "Rate" は 50 回の試行に対する学習の成功率を表す．ただし，問題 A に対してはどの学習法を用いた場合も学習時間にほとんど差が出なかったため記載していない．表 1 から，この問題に対しては MSPM がもっとも反復回数が少ないことがわかる．その他の実験結果でも，MSPM は反復回数，学習時間，学習の成功率の各々について on-line BP や BP モーメント法に勝ることが確かめられた．その他の実験結果および実験状況の詳細については当日発表する．

表 1: Results for test problems A and B

Prob.	Algorithm	Order	CPU sec	#Ite.	Rate
A	on-line BP	cyclic	-	9.40	50/50
		almost	-	3.04	50/50
	BP moment	cyclic	-	6.58	50/50
		almost	-	4.94	50/50
	MSPM	cyclic	-	2.84	50/50
		almost	-	2.82	50/50
B	on-line BP	cyclic	0.549	596.5	46/50
		almost	0.321	460.7	46/50
	BP moment	cyclic	0.319	377.1	47/50
		almost	0.308	356.7	46/50
	MSPM	cyclic	0.314	72.2	50/50
		almost	0.224	55.5	50/50

参考文献

[1] D.E. Rumelhart, G.E. Hinton, R.J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, Vol. 323, pp. 533 - 536 (1986)

[2] 茨木, 福島： 最適化の手法, 共立出版 (1993)

[3] 巽, 福島："逐次射影法としての誤差逆伝搬法," システム制御情報学会論文誌, Vol. 8, pp. 204-211 (1995)

[4] K. Tatsumi and M. Fukushima, "A Successive Projection Method for Binary Pattern Recognition with Multilayer Feedforward Neural Networks," *International Journal of Systems Science*, to appear

[5] 巽, 福島："階層型ニューラルネットワークに対する逐次射影法の改良," 1996 年日本 OR 学会春期研究発表会アブストラクト集, pp. 234-235

[6] Y. Shimazu, M. Fukushima and T. Ibaraki, "A Successive Over-Relaxation Method for Quadratic Programming Problems with Interval Constraints," *Journal of the Operations Research Society of Japan*, Vol. 36, No. 2, pp. 73-89 (1993)