改良型 online 記憶制限準ニュートン法によるニューラルネットワークの学習

阿部俊和　坂下善彦　二宮　洋

湘南工科大学工学部情報工学科 〒251-8511 神奈川県藤沢市辻堂西海岸 1-1-25

1．はじめに

ニューラルネットワークの学習問題を無制約最適化問題としてとらえることで，最急降下法に基づく勾配法を用いたバックプロパゲーション(BP)法が提案された．BP 法に対する研究の 1 つとして学習時のデータの与え方とネットワークの重みの更新に関する研究としてオンライン学習とバッチ学習がある．BP 法と同様に勾配法に基づく学習アルゴリズムに準ニュートン法がある．しかし，準ニュートン法は上記の区別ではバッチ学習となる．バッチ学習では，すべての，学習データセットをネットワークの入力した後に重みを更新するため，計算時間及びメモリ使用量が膨大となってしまう問題がある．このため，近年，online 準ニュートン法が提案された[1]．しかし，online 準ニュートン法をそのまま応用しただけでは，学習が困難な入出力パターンの学習を考えた場合，局所最適解から抜け出すことはできない．この問題を克服する為，改良型 online 準ニュートン法が提案された[2]．一方，準ニュートン法を大規模な問題に適用できるような工夫として，記憶制限準ニュートン法が注目されている．本研究では，記憶制限準ニュートン法を改良型 online 準ニュートン法に導入することを考える．これにより，改良型 online 準ニュートン法の収束性を損なうことなく，学習の計算時間及びメモリ使用量を削減することができることをするアルゴリズムを提案する．

2. ニューラルネットワーク

本研究では入出力関係として(1)の関係を持つニューラルネットワークを考える，

$$\mathbf{o}_p = f_{NN}(\mathbf{w}, \mathbf{x}_p) \qquad (1)$$

ここで，$\mathbf{x}_p$ は p 番目の入力ベクトルを，$\mathbf{o}_p$ は p 番目の入力ベクトルに対するニューラルネットワークの出力ベクトルを示す，$\mathbf{w}$ はニューラルネットワークの重みベクトルを示す．中間層のニューロンの入出力関係はシグモイド関数: $f(x) = 1/\left(1-e^{-x}\right)$ を持つとする.ここで，p 番目の入力ベクトル $\mathbf{x}_p$ に対する教師信号ベクトルを $\mathbf{d}_p$ とすると,誤差関数(2)が定義できる．

$$E(\mathbf{w}) = \frac{1}{2}\frac{1}{P_{T_r}}\sum_{p\in T_r}\left\|\mathbf{d}_p - \mathbf{o}_p\right\|^2 \qquad (2)$$

ここで，T_r は P_{T_r} 個の学習サンプルデータ $\left\{[\mathbf{x}_p, \mathbf{d}_p]\, and \;\; p \in T_r\right\}$ を持つ学習データセットである．学習ではネットワークの重みベクトル $\mathbf{w}$ に関して(2)を最小化する無制約最適化問題を考える．

3. 改良型 online 準ニュートン法

BFGS 公式に基づく準ニュートン法(BFGS)は無制約非線形計画法の中では最も有効な手法である．BFGS をより複雑な入出力特性を持つ問題の学習が可能なアルゴリズムへ改良したものに，改良型 online 準ニュートン法(ioBFGS)がある[2]．ioBFGS では，まず学習セット T_r を

$$T_r = \{T_{r\_1}, T_{r\_2}, \cdots, T_{r\_S}\} \qquad (3)$$

のように S 個の部分集合(ここではセグメントと呼ぶ)に分割する．このセグメント毎に重みを更新する手法が online BFGS がある。一方 ioBFGS では，このセグメントをオーバーラップさせながら増加させていくことで，重みの更新時の学習パターン数を増加させていく手法である．このオーバーラップを Cat 回行った場合の分割された学習パターンセット T_r は以下の様になる．

$$T_r = \left\{T_{r\_Cat(c,1)}, T_{r\_Cat(c,2)}, \cdots, T_{r\_Cat(c,S-1)}, T_{r\_Cat(c,S)}\right\} \qquad (4)$$

ここで，1 つの部分集合(カテゴリ)を

$$T_{r\_Cat\,(c,i)} = \left\{T_{r\_i}, \cdots, T_{r\_(c+i-1)\%S}\right\} \qquad (5)$$

とおいた．これに伴い,カテゴリ毎の誤差関数を(6)に示すように再定義する．

$$E_{Cat(c,i)}\left(\mathbf{w}, \left\{\mathbf{x}_{p\_Cat(c,i)}\right\}\right) = \frac{1}{2}\frac{1}{\dfrac{P_{T_r} * Cat(c,i)}{S}}\sum_{\substack{p\_Cat(c,i) \\ \in T_{r\_Cat(c,i)}}}\left\|\mathbf{d}_{p\_Cat(c,i)} - \mathbf{o}_{p\_Cat(c,i)}\right\| \qquad (6)$$

ioBFGS では(6)を用いて更新する．これにより，反復の始めは oBFGS となり，最終的には BFGS と同じ反復となる手法である． ioBFGS のアルゴリズムを以下に示す．

1. $k = 1$;
2. Initialize $\mathbf{H}^k$ by the unit matrix I;
3. $While(k < k_{\max}\, or \,\| E_{Cat(c,i)}\left(\mathbf{w}^k, \left\{\mathbf{x}_{p\_Cat(c,i)}\right\}\right) / \partial \mathbf{w}^k \| > \varepsilon_1)$

4. Calculate $\mathbf{g}^k = -\mathbf{H}^k \partial E_{Cat(c,i)}\left(\mathbf{w}^k, \{\mathbf{x}_{p\_Cat(c,i)}\}\right) / \partial \mathbf{w}^k$;
5. Execute Line search to decide the step size α^k;
6. Update $\mathbf{w}^{k+1} = \mathbf{w}^k + \alpha^k \mathbf{g}^k$;
7. Calculate $\mathbf{H}^{k+1}$ by using BFGS formula [2]
8. k + k+1;
9. Updata Cat;
10. *if* $Cat < S$ *then* $Cat = Cat + i$
    else $Cat = S$;
11. *EndWhile*
12. Re*turn* $\mathbf{w}^k$

4. 改良型 online 記憶制限準ニュートン法

準ニュートン法を大規模問題に適応できるような工夫として，記憶制限準ニュートン法(LBFGS)が提案されている[1]．通常の準ニュートン法では,近似行列を保存するのに大量の記憶容量とメモリ使用量を必要とするのに対して，近似行列を保存しないで m 本のベクトルに保存しておくことで，必要な時に随時ベクトルの内積計算をして探索方向を求めることができる．探索方向の計算手順のアルゴリズムを以下に示す．

1. $\mathbf{g}^k = E_{Cat(c,i)} / \partial \mathbf{w}^k$;
2. *for* $i := k-1, k-2, \cdots, k-m$
   (a) $\gamma^i = \mathbf{s}^{k-i^T} \mathbf{g}^k / \mathbf{s}^{k-i^T} \mathbf{y}^{k-i}$;
   (b) $\mathbf{g}^k = \mathbf{g}^k - \alpha_i \mathbf{y}^{m-i}$;
3. $\mathbf{g}^k = \left(\mathbf{g}^k / k\right) \sum_{i=1}^{k} \mathbf{s}^{k-i^T} \mathbf{y}^{k-i} / \mathbf{y}^{k-i^T} \mathbf{y}^{k-i}$;
4. *for* $i = k-m, k-m+1, \cdots, k-1$
   (a) $\beta = \mathbf{y}^{k-i^T} \mathbf{g}^k / \mathbf{y}^{k-i^T} \mathbf{s}^{k-i}$;
   (b) $\mathbf{g}^k = \mathbf{g}^k + (\gamma^i - \beta) \mathbf{s}^{k-i}$;
5. *Return* $\mathbf{g}^k$

上記のアルゴリズムを ioBFGS の step4 及び step7 と入れ替えることにより改良型 online 記憶制限準ニュートン法(ioLBFGS)のアルゴリズムとなる．

5. シミュレーション結果

本研究では，LBFGS，及び oLBFGS[1]に対する ioLBFGS の有効性をシミュレーションにより示す．例題としては，3 層ニューラルネットワークに対して(7)のモデル化を考える．

$$y(x, a, b) = 1 + (x + 2x^2)\sin(-ax^2 + b) \qquad (7)$$

ニューラルネットワークの入力は $x \in [-4,4]$，$a \in [-0.5, 0.5]$，$b \in [-0.5, 0.5]$ とし，出力は y とする．x は 0.1 刻みで学習データを求める．又，Ex1 では a を 0.02 刻みで学習データを求め，b は 0 に固定することで，学習データは 4000 となる．ネットワークの構造は，入力層，中間層及び出力層のニューロン数を 2，25，及び 1 とする．Ex2 では a を 0.05 刻みで，b を 0.2 刻みで学習データを求めることで，10080 パターンとなる．ネットワークの構造は，入力層，中間層及び出力層のニューロン数を 3，30，及び 1 とする．各例題に対して，シミュレーションを Ex.1 では異なる初期値を用いて 20 回、Ex.2 では 10 回実行し，その平均(Ave)と最良解(Best)を求めた．その結果を表 1 に示す．ここで，各シミュレーションの評価は[2]と同様に教師信号とのずれの割合として求めた．表より[2]と同様に ioLBFGS が LBFGS 及び oLBFGS と比較して，よい結果となった．これは，ioBFGS の収束性を損なうことなく最適解を見つけることができたことが分かる．

6. まとめ

本研究では，ニューラルネットワークの学習アルゴリズムとして改良型 online 準ニュートン法(io BFGS)に記憶制限準ニュートン法(LBFGS)を導入した改良型 online 記憶制限準ニュートン法(ioLBFGS)を提案した．例題を用いたコンピュータシミュレーションを行った結果，LBFGS,ioLBFGS ともに，Best は非常に小さな値を得ることができた．ioBFGS と同様に LBFGS 及び oLBFGS と比較して，ioLBFGS は有効な手法であることが明らかになった．

参考文献

[1] Nicol N. Schraudolph, Jin Yu, and Simon Gunter, “A stochastic quasi-Newton method for online convex optimization”, In Proc. 11th Intl.Conf. Artificial Intelligence and Statistics (AIstats), San Juan, Puerto Rico, March 2007.

[2] 二宮 洋：改良型 online 準ニュートン法によるニューラルネットワークの学習，信学技報,vol.109, no.269, NLP2009-115, pp.187-192, 2009 年 11 月

表 1 シミュレーション結果

Algorithm	Ex1				Ex2			
	Training error		Test error		Training error		Test error	
	Ave	Best	Ave	Best	Ave	Best	Ave	Best
LBFGS	6.20E+00	6.82E-01	6.20E+00	6.62E-01	8,71E+00	8.71E+00	8.65E+00	8.65E+00
oLBFGS	6.08E+00	5.15E+00	6.21E+00	5.21E+00	8.60E+00	8.57E+00	8.56E+00	8.53E+00
ioLBFGS	1.51E+00	7.48E-01	1.48E+00	7.31E-01	2.98E+00	1.49E+00	2.85E+00	1.39E+00