6R-6

勾配学習法に対する学習データの与え方の影響に関する研究

阿倍俊和† 坂下義彦‡ 二宮洋‡
湘南工科大学大学院工学研究科電気情報工学専攻 〒251-8511 藤沢市辻堂西海岸 1-1-25†
湘南工科大学工学部情報処理工学科 〒251-8511 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(NN)の学習問題に対する最も有効性な解法の１つに準ニュートン法に基づく勾配法を用いた学習アルゴリズムがある．この手法に関する研究の１つに，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着目したオンライン学習法及びバッチ学習に関する研究がある[1]．これらの手法の特徴としては以下の点が挙げられる．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，収束は速いが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．一方，オンライン学習の収束は遅いが，計算量が少なく局所解を抜け出す能力を有する．しかし，オンライン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かった．この問題を克服する為，改良型オンライン準ニュートン法が提案された[2]．しかしながら，オンライン学習及び改良型オンライン学習法では，学習データの与え方の順序は考慮に入れず，一定の順序に与えている．

本研究では，改良型オンライン準ニュートン法に対して学習データの与え方の順序に対してランダム性を付与することを考える．具体的には，反復毎に学習を与える順序を変化させ，収束性に対するその影響に関してシミュレーションにより検証する．

2. バッチ学習とオンライン学習

本研究では入出力関係として(1)の関係を持つ階層型ニューラルネットワーク(LNN)を考える．

$$\mathbf{o}_p = f_{\mathrm{NN}}\left(\mathbf{w}, \mathbf{x}_p\right) \tag{1}$$

ここで，$\mathbf{x}_p$はp番目の入力ベクトルを，$\mathbf{o}_p$は$\mathbf{x}_p$に対する LNN の出力ベクトルを示す．また，LNN の重みベクトルを$\mathbf{w}$とする．中間層のニューロンの入出力関係はシグモイド関数とする．ここで，$\mathbf{x}_p$に対する教師信号ベクトルを$\mathbf{d}_p$とすると，誤差関数は(2)と定義できる．

$$\mathrm{E}(\mathbf{w})=\frac{1}{\mathrm{P}_{\mathrm{T}_r}}\sum_{p\in\mathrm{T}_r}\mathrm{E}_p(\mathbf{w})=\frac{1}{\mathrm{P}_{\mathrm{T}_r}}\sum_{p\in\mathrm{T}_r}\frac{1}{2}\left\|\mathbf{d}_p\text{-}\mathbf{o}_p\right\|^2 \tag{2}$$

ここで，T_rは学習データセット$\{\mathbf{x}_p, \mathbf{d}_p \text{ and } p \in \mathrm{T}_r\}$を示し，データ数は$\mathrm{P}_{\mathrm{T}_r}$とする．学習は LNN の重み$\mathbf{w}$に関して(2)の勾配法を用いて解くことになる．ここで，バッチ学習では，1回の重み$\mathbf{w}$の更新に対して(2)の勾配を用いる．一方，オンライン学習では，全学習サンプルの部分集合毎の誤差関数の勾配を用いて重み$\mathbf{w}$を更新する．ここで，この部分集合をセグメントと呼ぶ．以下では，学習データセットT_rを複数個(Seg)のセグメント$\mathrm{T}_r = \{\mathrm{T}_{r_1}, \mathrm{T}_{r_2}, \cdots, \mathrm{T}_{r_Seg}\}$に均等に分割することを考える．

3. 学習データの与え方を考慮したオンライン準ニュートン法

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も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，改良型オンライン準ニュートン法(ioBFGS)がある[2]．ioBFGS では，学習のはじめはオンライン準ニュートン法(oBFGS[1])と同様に Seg 個のセグメントに分割($\mathrm{P}_{\mathrm{T}_{r\_s}} = \mathrm{P}_{\mathrm{T}_r}/Seg$)し，セグメント毎に重みを更新する．次に，1 つのセグメント内の学習データ数を増加させ，oBFGS と同様の反復を行う．この操作を続けていくことで，ioBFGS は 1 つのセグメント内の学習データ数が徐々に増加し，最終的にはバッチ学習(BFGS)となる．セグメントを増加させる手法には２つある．１つは，１つのセグメント内のデータ数を単純に 2 倍，3 倍…と増加させる手法である(ioBFGS1)．2 つ目は，降り合うセグメントを重複させながら，データ数を増加させる手法である(ioBFGS2)．この時，c 回の操作を行った場合の学習データセットT_rはそれぞれ，

$$\mathrm{T}_r = \left\{\mathrm{T}^{\mathrm{ioBFGS1}}_{r\_Grp(c,1)}, \cdots, \mathrm{T}^{\mathrm{ioBFGS1}}_{r\_Grp(c,Seg/c)}\right\} \tag{3}$$

$$\mathrm{T}_r = \left\{\mathrm{T}^{\mathrm{ioBFGS2}}_{r\_Grp(c,1)}, \mathrm{T}^{\mathrm{ioBFGS2}}_{r\_Grp(c,2)}, \cdots, \mathrm{T}^{\mathrm{ioBFGS2}}_{r\_Grp(c,Seg)}\right\} \tag{4}$$

A Study on Effect of Feeding Method of Training Data in Gradient based Training Algortihm
Shonan Institute of Technology†‡
Toshikazu Abe† Yoshihiko Sakashita‡ and Hiroshi Ninomiya‡

となる．ここで，ioBFGS2における c 個のセグメントを含む i 番目の部分集合

$$\mathrm{T}_{r\_Grp(c,i)}^{\mathrm{ioBFGS2}} = \left\{\mathrm{T}_{r\_j\%(Seg+1)} : j = i, \cdots, i + c - 1\right\} \quad (5)$$

となる．

本研究では，学習データの与え方の順序にランダム性を付与することを考える．ここで，ランダム性を付与しない手法を R0 とする．ランダム性は具体的には次の 2 つの手法を考える．1 つ目は学習データ毎の順序をランダムに交換させる手法(R1)である．2 つ目はセグメント毎の順序をランダムに交換させる手法(R2)である．これらの交換をセグメント内の学習データ数を増加させる毎に行うことで，学習データの与える順序にランダム性を持たせる. 上記の手法を oBFGS, ioBFGS1 及び ioBFGS2 に適用し, その収束をシミュレーションを用いて検証する.

4.シミュレーション

本章では 3 層の LNN(3LNN)を用いて oBFGS, ioBFGS1 及び ioBFGS2 に対して学習データの与え方をランダムにする手法を適用し，有効性をシミュレーションにより示す．例題には，関数近似問題のベンチマーク(6)及び(7)を用いる[2][3][4].

$$\mathrm{y}(x_1, x_2) = 1.9(1.35 + \mathrm{e}^{x_1}\mathrm{e}^{-x_2} \sin(13(x_1 - 0.6)^2)\sin(7x_2)) \quad (6)$$

$$\mathrm{y}(x, a) = 1 + (x + 2x^2)\sin(ax^2) \quad (7)$$

ここで，(6)に対しては$0 \le \mathrm{x_i} \le 1, \forall \mathrm{i}$として学習データを与える．従って，(6)の学習データ数$\mathrm{P}_{\mathrm{T}_r}$は 440 となり，中間層のニューロン数($\mathbf{H}$)を 20 とした．(7)に対しては入力 a の与え方により，2 種類の例題 Ex.(7)-1 及び-2 を考える．Ex.(7)-1 は$a = 0$とする. 従って, 学習データ数は$\mathrm{P}_{\mathrm{T}_r} = 400$となり，$\mathbf{H} = 7$とした．Ex.(7)-2 は$|a| \le 1$とする．これにより，$\mathrm{P}_{\mathrm{T}_r} = 1680$となり，$\mathbf{H} = 45$とした．各手法は 100 回のシミュレーションを実行し，(2)に示す学習誤差の平均値(Ave$\times 10^4$)及び最小値(Best$\times 10^4$)を用いて評価する．

シミュレーション結果を表 1 に示す．表 1 より，学習データを増加させていく改良型オンライン学習(ioBFGS1 及び ioBFGS2)はオンライン学習(oBFGS)と比較して，最も小さな誤差を得られていることが分かる．これは，セグメントを増加させていくことで，局所最適解から抜け出す能力があると考えられる．オンライン学習は，$\mathrm{P}_{\mathrm{T}_{\mathrm{r_s}}}$の値によらず，すべての例題に対して局所最適解から抜け出すことはできないことが分かる．次に，学習データにランダム性を持たせた学習(R1 及び R2)と R0 を比較する．Ex.(6)では，平均値及び最小値ともに R1 及び R2 と R0 を比較して，あまり差がないことが分かる．しかし，複雑な非線形特性を持つ Ex.(7)-1 及び Ex.(7)-2 では，最小値は差がないが, 平均値は R1 や R2 の方が良い結果となった．これは，ランダム性が局所最適解から抜け出すことに少なからず影響を与えていることを示していると考えられる.

表 1 シミュレーション結果

Ex		$\mathrm{P}_{\mathrm{T}_{\mathrm{r\_s}}}$	oBFGS	ioBFGS1	ioBFGS2
			Ave/Best	Ave/Best	Ave/Best
(6)	R0	10	79.6/47.1	4.91/3.00	4.41/2.84
		88	58.3/29.3	4.65/2.85	4.61/3.05
	R1	10	220.2/64.8	5.17/3.17	4.55/2.95
		88	45.6/23.8	5.25/3.06	4.73/2.84
	R2	10	113.9/53.2	4.78/3.02	4.40/2.93
		88	68.9/36.0	4.69/2.65	4.87/2.81
(7)-1	R0	10	571.1/542.5	36.4/1.86	30.3/1.86
		80	441.1/96.9	35.5/1.86	33.0/1.86
	R1	10	584.6/568.5	23.8/1.86	24.1/1.86
		80	210.5/79.3	26.7/1.86	22.9/1.86
	R2	10	556.1/535.9	28.2/1.86	23.4/1.86
		80	513.9/376.6	34.1/1.86	33.0/1.86
(7)-2	R0	10	384.0/365.3	16.5/2.91	15.5/2.69
	R1	10	384.5/383.0	17.1/2.58	6.06/1.86
	R2	10	383.7/341.8	15.9/3.34	7.69/1.71

5. 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オンライン準ニュートン法の学習データの与え方に関する考察を行った具体的には学習データの与える順序にランダム性を持たせ，収束性に対する影響をシミュレーションにより検証した．シミュレーション結果からランダム性が局所最適解から抜け出すことに少なからず影響を与えていることが分かる

参考文献

[1] N. N. Schraudolph, J. Yu, and S. 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 オンライン 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 A 論, vol.593-A, no.12, pp.828-832, 2010 年 12 月

[3] 阿部 俊和, 坂下 善彦 二宮 洋, "学習データの与え方を改良した勾配学習アルゴリズムを用いたニューラルネットワークの学習に関する研究", 信学技報, vol.110, no.82, NLP2010-9, pp.57-62, 2010 年 6 月

[4] 阿部 俊和, 坂下 善彦 二宮 洋, "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", 情報処理学会創立50周年記念全国大会, pp.2-447-2-448 2010 年 3 月