

ニューラルネットワークの学習における非線形最適化手法の評価検討

阪上, 茂生
松下電器産業株式会社中央研究所

高木, 英行
松下電器産業株式会社中央研究所

香田, 敏行
松下電器産業株式会社中央研究所

久木, 泰治
松下電器産業株式会社中央研究所

他

<https://hdl.handle.net/2324/4481553>

出版情報 : pp. 7-307-7-308, 1989. 電子情報通信学会
バージョン :
権利関係 : Copyright (C) IEICE



SD-1-13 ニューラルネットワークの学習における
非線形最適化法の評価検討Evaluation of Non-linear Optimization Methods
Applied to Learning of Artificial Neural Network阪上茂生^{*1} 高木英行^{*1} 香田敏行^{*1} 木泰治^{*1} 戸川隼人^{*2}
S. Sakaue H. Takagi T. Kouda Y. Shimeki H. Togawa^{*1}: 松下電器産業(株) 中央研究所 ^{*2}: 日本大学理工学部

Matsushita Electric Ind. Co. Ltd., Central Research Lab. Nihon University, College of Science and Technology

1. まえがき

従来ニューラルネットワークの学習法として、誤差逆伝搬法があるが、学習時間がかかりすぎるといえる課題がある。文字認識や音声認識などの実際問題にニューラルネットワークを適用しようとする、その学習に数時間以上かかるのが普通である。本稿では、誤差逆伝搬法に非線形問題の高速解法であるDSCサーチおよびFletcher-Reevesの方法を導入し、その有効性を評価した。学習の適用問題は、EOR、binary decoder、binary encoder及びsine曲線による平面分割とした。

2. 誤差逆伝搬法¹⁾

階層構造を持つニューラルネットワークの学習法である誤差逆伝搬法は、下記(1)式の誤差Eの最小化問題に最急降下法を適用したものである。誤差Eは

$$E = E(\vec{w}) \dots (1)$$

と表される。ここに $\vec{w}$ は重み $w[i, j]$ を要素とする列ベクトルである。重み $w[i, j]$ の変化量は

$$\Delta w[i, j] = -\epsilon * (\partial E / \partial w[i, j]) + \alpha * \Delta w_{old}[i, j] \dots (2)$$

で求められる。ここで ϵ は学習パラメータ、 α は加速パラメータ、 $\Delta w_{old}[i, j]$ は前回の重み $w[i, j]$ の変化量である。

3. 高速学習アルゴリズム

3-1. DSCサーチ²⁾

DSCサーチは、最急降下法の高速化アルゴリズムであり、最急降下方向に対する最適な学習パラメータを設定しながら学習を進める方法である。最急降下の方向即ち

$$\Delta \vec{w}_{dir} = \left[\partial E / \partial w[i, j] \right] \dots (3)$$

を重み空間における誤差Eの最小点の探索方向とし、 $\Delta \vec{w}_{dir}$ の方向における誤差Eの最小点を求める。この繰り返しにより誤差Eを最小化する。

3-2. Fletcher-Reevesの方法²⁾³⁾

Fletcher-Reevesの方法は、共役勾配法の一様であり、最急降下の方向とは異なる共役勾配の方向を誤差の最小値探索の方向とする方法である。重み空間における誤差Eの最小点の探索方向を

$$\Delta \vec{w}_{dir} = \left[\partial E / \partial w[i, j] \right] + \beta * \Delta \vec{w}_{old} \dots (4)$$

$$\beta = \frac{\sum (\partial E / \partial w[i, j])^2}{\sum (\partial E_{old} / \partial w[i, j])^2}$$

によって決定する。ここで、 $\Delta \vec{w}_{old}$ は前回の探索方向であり、 $\partial E_{old} / \partial w[i, j]$ は前回の探索における最急降下の方向である。

4. 評価実験

3層のニューラルネットワークの学習速度について、EOR、binary decoder及びbinary encoder、sine曲線による平面分割をその適用問題とし、従来の誤差逆伝搬法、DSCサーチ及びFletcher-Reevesの方法の学習速度を比較評価した。binary decoder及びbinary encoderについては適用問題の規模の変化に対する学習速度の変化を調べ、sine曲線による平面分割については、適用問題の非線形性に対する学習速度の変化を調べた。

実験条件

計算機 Sun3/60

閾値関数 $\tan^{-1}$ 誤差 $0.5 / NR / NP * \sum_p \sum_k (z[k] - t[k])^2$

NR: 出力層のユニット数

NP: 教師データのパターン数

 $z[k]$: 出力値 $t[k]$: 教師データ

	EOR	binary decoder			binary encoder			sine曲線 による 平面分割
		1bit	2bit	4bit	1bit	2bit	4bit	
入力層のユニット数	2	1	2	4	2	4	16	2
中間層のユニット数	2	2	4	16	2	4	16	4
出力層のユニット数	1	2	4	16	1	2	4	1
総リンク数	9	10	32	352	9	30	340	17
教師データのパターン数	4	2	4	16	2	4	16	42

4-1. EORの学習に関する検討

EO Rゲートを3層のニューラルネットワークで構成する際に、重みの初期値を10通りに変化した時の学習速度について評価した結果を図1に示す。10通りの重みの初期値のうち2通りについては、従来法でもDSCサーチやFletcher-Reevesの方法でも収束しなかった。

従来法では、これらも含めて4通りの初期値で収束しなかった。

4-2. 適用問題の規模に関する検討

--- binary decoder 及び binary encoder

binary decoder及びbinary encoderを1bit, 2bit 及び4bit で構成し、適用問題の規模が大きくなった時、各学習方法で学習速度がどのように変化するかを評価した結果を図2及び図3に示す。

4-3. 適用問題の非線形性に関する検討

— sine曲線による平面分割

$y = A * \sin(\pi * x)$ ($-1 \leq x \leq 1$)
で(x, y)平面を分割する問題を考える。sine曲線の振幅Aが大きくなるほど適用問題の非線形性が強くなる。

教師データとしては

$$y = A * \sin(\pi * x) + 0.2$$

$$(-1 \leq x \leq 1 \text{ step } 0.1)$$

$$y = A * \sin(\pi * x) - 0.2$$

$$(-1 \leq x \leq 1 \text{ step } 0.1)$$

を与える。振幅Aを0, 0.2, 0.4と変化した時、各学習方法で学習速度がどのように変化するかを評価した結果を図4に示す。

5. まとめ

- (1) 収束条件を誤差 < EPSILONとすると
 - ・ EPSILON > $5 * 10^{-3}$ では高速解法の効果はない
 - ・ EPSILON < $5 * 10^{-5}$ では高速解法により学習時間が1/10以下になる。
- (2) DSCサーチとFletcher-Reevesの方法とは同程度の効果がある。

今後はパターン認識等の実際的な適用問題について、これら高速解法を導入したい。

6. 謝辞

日頃御指導頂く新田所長、松島室長に謝意を表します。

参考文献

- 1) Rummelhart, D. E., Hilton G. E., Williams, R. J., "Learning representations by back-propagating errors", Nature Vol. 323 NO.9 Oct. 1986
- 2) 高木他, "非線形最適化手法を用いたニューラルネットワーク学習アルゴリズムの高速化", 電子情報通信学会春季全国大会(1989)
- 3) コワリック, オスボーン著, 山本、小山訳, "非線形最適化問題", 培風館(1970)

— 従来法
— DSCサーチ
- - - Fletcher-Reevesの方法

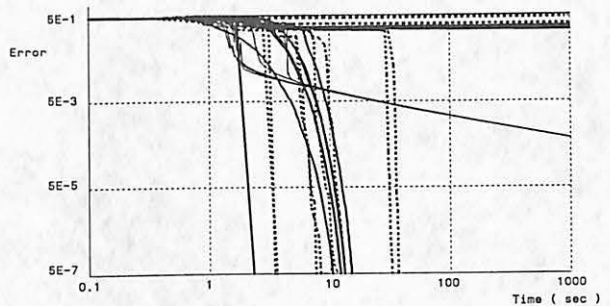


図1 EORの学習時の誤差の時間変化

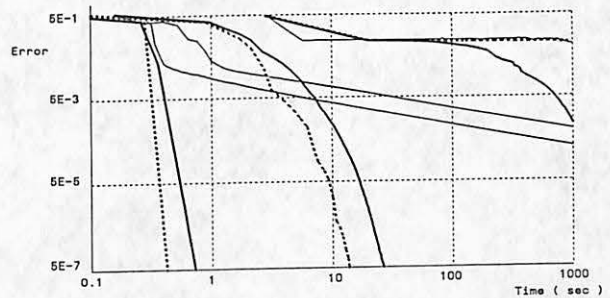


図2 binary decoderの学習時の誤差の時間変化

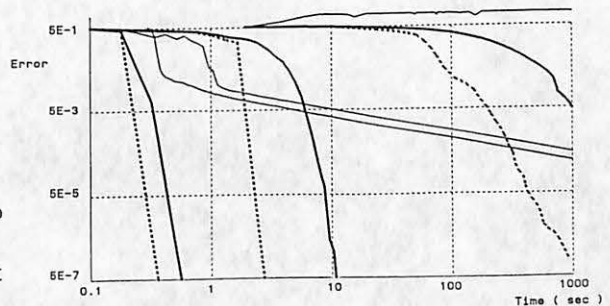


図3 binary encoderの学習時の誤差の時間変化

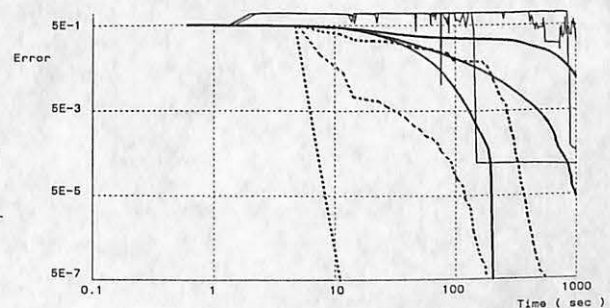


図4 sine曲線による平面分割の学習時の誤差の時間変化