1995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

非線形方程式による階層型ニューラルネットワークの学習アルゴリズム

広島県立大学　青木兼一　AOKI kenichi
広島県立大学　*藤原輝明　FUJIWARA teruaki

1．階層型ニューラルネットワークの学習アルゴリズム

階層型ニューラルネットワークは神経回路を単純に模擬したネットワークで、このネットワークは優れた学習効果をあげることが報告されており、種々の分野に適用されている。

このネットワークの学習アルゴリズムとは、各パターンPへの入力ベクトルI_pに対し、教師信号ベクトルT_pに等しい出力ベクトルO_pを満足させる結合荷重ベクトルW、しきい値ベクトルθを調整する問題である。

このアルゴリズムとしては各パターン、各出力に対して教師信号との誤差の二乗和を最小にする方法がとられており、その代表的な方法が BP 法[1]である。

このBP法は基本的にグレジエント法を用いているので収束性が悪く、この収束性を改善するために無制約最適化の方法である共役傾斜法、疑似ニュートン法などが用いられている。

階層的ニューラルネットワークは構造的には単純であるが、神経回路の入出力特性のステップ関数を微分可能にするため、入出力特性にシグモイド関数を用いている。教師信号が0,1の値をとる場合は非線形性が強くなり、収束しなかったり極小値にトラップする場合があり、非線形最適化の問題としては悪条件の問題である。

このため教師信号が 0, 1 の場合はこれを$\varepsilon, (1-\varepsilon)$, $\varepsilon > 0$としており、εをできるだけ小さくすることが収束性の問題である。

BP 法は各パターンに対し誤差を評価して変数W,θを調整する方法で、これを逐次修正法という。これに対し、すべてのパターンに対する誤差を評価して変数を調整する方法は一括修正法といわれている。

現在大規模なニューラルネットワークのハードウェアが開発されていないが、このようなハードが開発されるならば逐次修正法の効果は上がるものと考えられる。逐次修正法では収束を良くするために慣性項といわれる過去のパターンの履歴による補正を行っているが、この根拠は明確でない。

2．非線形方程式を用いたアルゴリズム

1で説明したpパターン、j出力は、W,θの関数として$O_{pj}(W,\theta)$と表される。従ってO_{pj}を教師信号T_{pj}に一致させるためには、(1)式が成立すればよい。

$$O_{pj}(W,\theta) = T_{pj}$$
$$(p = 1,\cdots,P\text{、}j = 1,\cdots,J) \qquad \cdots(1)$$

(1)式は非線形方程式であるからこれを(2)式のように表す。

$$f(x) = 0,\ f \in R^m,\ x \in R^n \qquad \cdots(2)$$

ニューラルネットワークの学習では$m \leq n$である。

筆者は従来の誤差二乗関数の最小化ではなく、学習アルゴリズムを(2)式の非線形方程式の解を求める計算方法と考え、$m < n$なる問題のニュー

トン法の計算方法を作成し、この計算方法の収束性を高めるためにホモトピー法を併用したアルゴリズムを作成した[2]。

このアルゴリズムではホモトピー法のきざみを小さくとれば、スラップ関数に近い学習問題の解を求めることができた。本文では(2)式の非線形方程式の解法に、

(i) LQ 分解を用いて、とくに線形近似 $A\Delta x = b$, A : $m - m$行列 のランク落ちに対して PAQ なる最小二乗和形式を導入している。

(ii)シグモイド関数 $y(x)$ の微分は $y' = y(1-y)$ となり、Jacobian 行列 A はすべて $y(1-y)$ を含むため、出力が 0 または 1 に近くなるときは A 行列が零行列に近くなる。これを防ぐため、$b / y(1-y)$ とするスケール変更を行っている。

(iii)逐次修正法は非線形方程式の解法である Brent 法[3]と全く同じであり、パターンごとにこの方法を用いることにより慣性項の概念を明確にしている。

3．アルゴリズム

(i)LQ 分解

(2)式の線形近似を(3)式で表す。

$$Ax = b,\ A: m - n \text{行列},\ m < n \quad \cdots (3)$$

A は Jacobian 行列を表す。

A のランクが m の場合は(4)式が成立する。

$$AQQ^t x = \left[\ L\ \right]\left[\ O\ \right] P = b \quad \cdots (4)$$

ここに Q は直交行列、L は $m - m$ 下三角行列、$P = Q^t x$ である。A のランクが $r < m$ の場合は(4)式の解が存在せず、(5)式が成立する。

$$PAQ = \begin{array}{c} \\ r \end{array}\!\!\begin{array}{c} r \\ \left[\begin{array}{c} L \\ \hline O \end{array}\right] \end{array}\left[\ O\ \right] \quad \cdots (5)$$

ここに P は直交行列。(5)式の表現は(2)式の解が存在しない場合、最小二乗法（誤差二乗関数）の意味で解を求める問題となる。

(ii)スケール変換

(3)式の b の要素は $T_{pj} - O_{pj}$ を表す。A 行列の pj 行は $(1-O_{pj})O_{pj}a_{pjk}$ となる。ここに a_{pjk} は、k 列要素である。

いま $O_{pj} = 0$ or 1 とすると、pj 行の要素はすべて零となる。これが LQ 分解を困難にする。しかし、$(T_{pj} - O_{pj}) / ((1-O_{pj})O_{pj})$ とスケール変換すれば、A 行列のある行が零ベクトルとなることを防ぐことができる。

(iii)逐次修正法

パターンごとの逐次修正法は、(1)式の J 個の非線形方程式の追加条件を考慮した Brent 法[3]を行うことである。従ってランク落ちのない場合は、文献[3]と同じ方法で慣性項を考慮した計算が行われる。

4．考察

以上のアルゴリズムは、1 で述べた ε を考慮した非線形不等式の解法[4]に拡張することができる。今後は本アルゴリズムによる汎化能力について検証を行いたい。

参考文献

[1] D.E.Rumelhart & J.L.McCelland:「Parallel Distributed Processing」(1988) MIT Press

[2] 青木,金指:「非線形方程式の求解によるニューラルネットの学習法」電気学会論文誌 C No3 (1993),pp.219-225

[3] J.J.More,M.Y.Cosnard:「Numerical Solution of Nonlinear Equations」ACM Transaction on Mathematical Software. Vol.5,No.1(1971)

[4] 巽,福島:「逐次射影法としての誤差逆伝搬法」OR 学会アブストラクト(1994),pp.71-72