4N－6

高次元アルゴリズムによる効率的な大局解探索法：ニューラルネットワークの学習問題への適用

河野 芳江　　下川 信祐　　新上 和正
（株）ＡＴＲ環境適応通信研究所

１．はじめに

発見学的最適解探索法の１つである高次元アルゴリズム（ＨＡ）[1]において、簡便に混合性を導入する一つの方法と混合性導入の効果について検討する。

ＨＡは他の最適解探索法と比べ、局所解に捕まりにくい、変数の多い問題に有効、プログラミングが簡単、などの特徴をもつと期待される方法である。ＨＡを効率的に機能させるためには、与えられた問題に対し、適切な混合性をもつハミルトン関数を設定することが重要である[2]。系が混合性をもつ場合、変数の組は解でない領域をすばやく通り過ぎ、解に相当する領域に長く滞在する性質があるため、これを利用して解を効率的に探すことができる。一方、混合性が十分でない場合には、その性質が失われ解を探すことは困難である。そこで、今回、極めて簡便に混合性を導入する方法を検討したので報告する。

具体例として、現時点では解くのが相当に困難である、多数の結合強度を最適設計しなければならないニューラルネットワークの学習問題[3]をとりあげ、本手法を適用した計算結果を紹介する。

２．混合性の簡便な導入法

Ｌ個の変数の組 $\mathbf{q}=(q_1,..,q_L)$ と、その変数に依存する評価関数 $V(\mathbf{q})$ が与えられたとする。ＨＡを用いて、評価関数 $V(\mathbf{q})$ を最小にする変数の組 $\mathbf{q}$ を見つける最適化問題を解くことを考える。$\mathbf{q}$（一般座標）の各成分に正準共役な変数（一般運動量）の組を $\mathbf{p}=(p_1,..,p_L)$ とする。ハミルトン関数は全微分可能であれば任意にその関数形を設定できるが、ポテンシャルエネルギー項に対応する評価関数は問題で与えられているため、人為的に調節できるのは主として運動エネルギー項である。ここでは、次のハミルトン関数を採用する。

$$H[\mathbf{q};\mathbf{p}]=\frac{1}{2}(\mathbf{A}\,\mathbf{p},\mathbf{p})+V(\mathbf{q}) \qquad (1)$$

運動エネルギーを常に非負とするために $\mathbf{A}$ はＬ次正値対称行列、$\det\mathbf{A}\neq 0$ を満たすとする。$\mathbf{A}$ として単純にＬ次単位行列 $\mathbf{E}_L$ を用いると、一般に、混合性が十分でない場合が多い。そこで、$\mathbf{A}$ としてＭ次対称行列 $\mathbf{b}$ を対角にＮ個並べた（Ｌ＝ＭＮ）、ブロック対角行列 $\mathbf{B}$ を用いることを提案する。

$$\mathbf{B}=\begin{bmatrix}\mathbf{b}_1 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{b}_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}\\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{b}_N\end{bmatrix} \qquad (2-1)$$

$$\mathbf{b}_i=\mathbf{S}\,\mathbf{D}_i\mathbf{S} \quad (i=1,\ldots,N) \qquad (2-2)$$

ここで、$\mathbf{D}_i$ は各対角要素が正のランダムな値をもつＭ次対角行列、

$$\mathbf{D}_i=\begin{bmatrix}\mu_1^i & 0 & \cdots & 0\\ 0 & \mu_2^i & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \mu_M^i\end{bmatrix} \quad (i=1,\ldots,N) \qquad (2-3)$$

$\mathbf{S}$ は次式で表されるＭ次対称行列とする。

$$\mathbf{S}=\mathbf{E}_M-c\begin{bmatrix}0 & 1 & \cdots & 1\\ 1 & 0 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 1\\ 1 & \cdots & 1 & 0\end{bmatrix} \qquad (2-4a)$$

$\mathbf{E}_M$ はＭ次単位行列、ｃは任意の定数で混合性の大小を左右するパラメータとなる。ただし、$\det\mathbf{S}\neq 0$ の条件より、次式を満たす必要があることに注意。

$$c\neq -1,\ \frac{1}{M-1} \qquad (2-4b)$$

以上のとおり設定した行列 $\mathbf{B}$ は正値対称、$\det\mathbf{B}\neq 0$ である。このハミルトン関数を用いて、初めに変数の組 $(\mathbf{q},\mathbf{p})$ を適当なランダムな初期値におき、正準

Effective global minimum search by Hamilton algorithm
:Application to learning in neural networks
KOHNO Yoshie, SHIMOGAWA Shinsuke, SHINJO Kazumasa
ATR　Adaptive Communications Research Laboratories

方程式

$$\frac{d\mathrm{q}_i}{dt}=\frac{\partial \mathrm{H}[\mathbf{q};\mathbf{p}]}{\partial \mathrm{p}_i},\quad \frac{d\mathrm{p}_i}{dt}=-\frac{\partial \mathrm{H}[\mathbf{q};\mathbf{p}]}{\partial \mathrm{q}_i}\quad (i=1,\dots,\mathrm{L}) \qquad (3)$$

にしたがって変数を運動させて解を探せばよい。混合性導入により新たに現れたパラメータ(Mやc等)は、試行計算で適当な値に設定する必要がある。

3．混合性導入の効果

適用例として、文献[3]のモデルを用いる。ユニット間の各結合が2つの結合強度をもち、分数型結合をする3層の階層型ネットワークモデルである。教師信号の入力と出力は線形関係にある。入力層、中間層、出力層のユニット数が、各々、20、20、24、教師信号数が 30 の場合に対して、上記の手法を適用した。

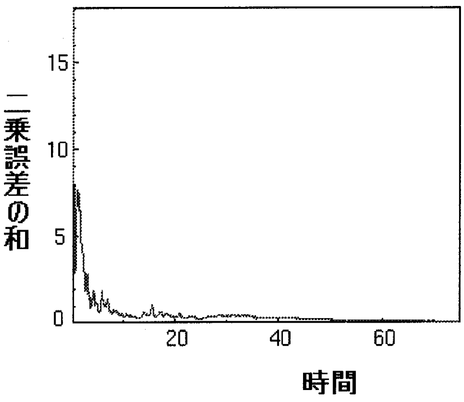

図1(a)混合性あり(A=B)

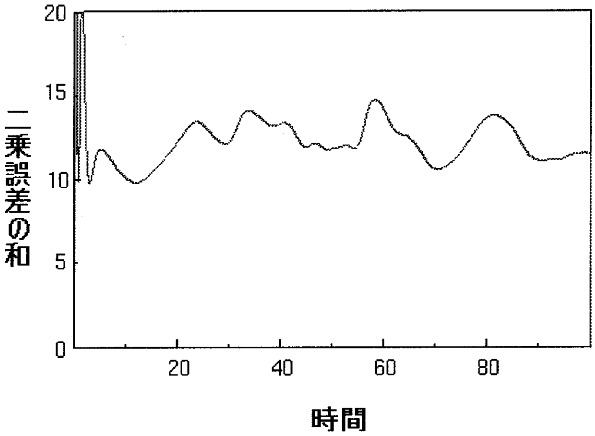

図1(b)混合性なし(A=E)

図1は、各々、(a)**A**＝**B** および (b)**A**＝**E**の場合に対する、出力と教師信号との二乗誤差の和の様子を時間（学習回数と比例）の関数として示したものである。(b)**A**＝**E**の場合には、混合性が十分ではないため、結局、解を見つけることができなかったが、混合性を増強した(a)**A**＝**B**の場合には効率的に解探索が行われている。

図2に、(a)**A**＝**B**の場合に対する、結合強度の一つであるω^1_{11}とその速度の運動の軌跡

($\omega^1_{11}(t)$, $d\omega^1_{11}(t)/dt$) を示した。解付近に運動が集中している様子が分かる。

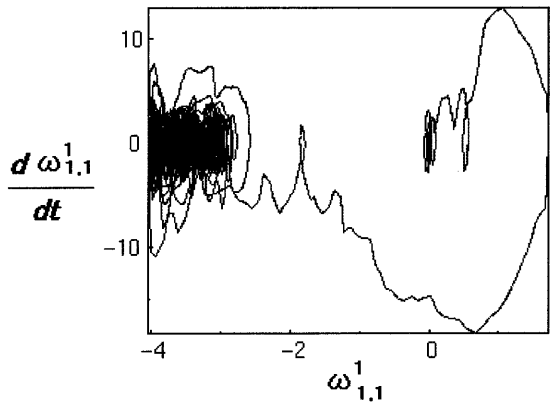

図2

4．おわりに

HAにおける混合性の簡単な導入法と混合性導入の効果ついて検討した。ニューラルネットワークの学習問題への適用例においては、混合性導入によりHAで効率よく学習が行われることが示された。今後はさらに、より良い混合性の導入法（適切なパラメータ値の設定法、別方法）や評価関数の工夫などを行い、さらに高速で安定した計算手法を検討してゆく予定である。

【参考文献】

[1] 新上："高次元アルゴリズム"、bit Vol.31,NO.7,p.2 (1999).

[2] 倉持,新上,下川："高次元アルゴリズムの高速化に関する検討"、信学会総合大会 A-1-9,p.9 (1999).

[3] 倉持,新上,下川,河野："拡張型結合をもつニューラルネットワークのHA学習"、情報処理学会第61回全国大会 4N-3 発表予定(2000).