高次演算処理ユニットをもつニューラルネットワークのＨＡ学習

１Ｐ－２

河野 芳江　　下川 信祐
（株）ＡＴＲ環境適応通信研究所

1．はじめに

パーセプトロン型ニューラルネットワーク（ＮＮ）の学習の高速化・大規模化を狙い、局所解に捕まりにくく多変数の問題に対して有利な最適化手法である高次元アルゴリズム（ＨＡ）を学習法とする手法の研究を進めている。これまで幾つかのテスト関数に対して本手法の連続関数近似能力を評価する計算機実験を行い[1,2]、その有効性を確認している。

学習高速化の最大のポイントの１つに、計算量の大幅な減少につながる中間ユニット数の削減がある。ユニット演算処理方法として、通常は入力と結合重みの積和演算が用いられるが、適切な高次演算の導入により、機能実現に要する中間ユニット数の削減が期待できる場合が多いと考えられる。直感的に意味を取りやすい演算を採用すれば、得られた結論の解釈が容易になるという利点もある。

そこで、今回、本手法において、中間ユニットの入力として、(a)超平面識別に対応する通常の積和演算を用いる場合、(b)超球面識別に対応する高次演算を用いる場合、の２つのケースを試み、用いる識別曲面の違いがどのように学習効率に影響するのかを比較・検討した。アフィン関数と正弦関数をテスト関数として計算機実験を行った結果を報告する。

2．モデル

３層パーセプトロン型ＮＮを考える。入力層、中間層、出力層のユニット数を各々n_1、n_2、n_3、入力－中間および中間－出力ユニット間の各結合重みを各々ω_{ij}とΩ_{jk}、中間層および出力層の各ユニットの閾値を各々θ_jとΘ_kとする。いま、n_s個の教師信号が与えられているものとし、s番目の教師信号の入・出力値をX^s_i, Y^s_kと表す。

A Method of Learning by Hamiltonian Algorithm in a Neural Network with Higher Order Operation Units
KOHNO Yoshie, SHIMOGAWA Shinsuke
ATR Adaptive Communications Research Laboratories

３層ＮＮの場合、主に中間層が入力空間における特徴領域抽出の役割を担う。j番目の中間ユニットの入力を$\eta_j(x_i;\omega_{ij},\theta_j)$とすると、これにシグモイド関数を出力関数として適用することにより、方程式$\eta_j=0$を満たす識別曲面を境に、単調でなだらかな遷移領域を経て出力が0と1の領域に分けられる。１つの中間ユニットは１つの識別曲面を指定し、n_2個の中間ユニットが各々指定する識別曲面の組み合わせですべての特徴領域が表現される。ユニット入力として通常用いられる積和演算

$$\eta^s_j = -\theta_j + \sum_{i=1}^{n_1} X^s_i \cdot \omega_{ij} \tag{1.a}$$

を用いれば識別曲面は超平面となる。閉じた特徴領域がある場合には、一般に、その表現のために多数の超平面が必要となり、その傾向は入力次元の増加とともに強くなる。閉領域の表現には閉領域を識別曲面とする演算の利用が有効だと考えられ、その単純な方法の１つに超球面識別がある。今回は、中心$(\omega_{1j},\omega_{2j},..,\omega_{n1j})$、半径$|\theta_j|$の超球面による識別に対応する次の高次演算

$$\eta^s_j = -\theta_j^2 + \sum_{i=1}^{n_1} (X^s_i - \omega_{ij})^2 \tag{1.b}$$

を用い、(1.a)の場合との比較を行った。いずれの場合も、１つの中間ユニットがもつ変数（ω_{ij}とθ_j）の数は同じである。中間ユニットの出力は、

$$\xi^s_j = f(\eta^s_j) \equiv \frac{1}{1+\exp(-\eta^s_j/T)} \tag{2}$$

で与える。

出力ユニットの入力には積和演算をとり、入力値をそのまま出力することとする。

$$\zeta^s_k = -\Theta_k + \sum_{j=1}^{n_2} \xi^s_j \cdot \Omega_{jk} \tag{3}$$

この出力層の役割は、教師信号の入力点におけるＮＮの実際の出力が教師信号出力値と等しくなるよう、

各中間ユニットの寄与を調整することにある。

コスト関数を、教師信号出力値と実際の出力の二乗誤差の和

$$V(\omega_{ij},\theta_j,\Omega_{jk},\Theta_k)=\sum_{s=1}^{n_s}\sum_{k=1}^{n_3}\left(\zeta_k^s-Y_k^s\right)^2 \qquad (4)$$

とし、これを最小にする結合重みと閾値をＨＡを用いて求める。

３．計算機実験結果

(a)超平面識別と(b)超球面識別の２つ場合に対し、各々、次の（ア）～（エ）のテスト関数に対する小規模なＮＮでの計算機実験を試み、関数依存性と入力次元依存性を比較した。

■１入力(ｘ)１出力(ｙ)；$0<x<1$

（ア） $y=0.67x+0.21$； $n_s=10$

（イ） $y=0.5+0.5\sin 2\pi x$； $n_s=50$

■２入力(x_1, x_2)１出力(ｙ)；$0<x_1, x_2<1$

（ウ） $y=-0.76x_1+0.19x_2+0.78$； $n_s=30$

（エ） $y=(\sin\pi x_1)(\sin\pi x_2)$； $n_s=200$

（ア）と（ウ）は１入力および２入力のアフィン関数で、各々、直線と平面を表し、単調で閉領域をもたない。一方、(イ) は定義域に１つの山と１つの谷をもつ１入力正弦関数、(エ)は１つの山をもつ２入力正弦関数であり、いずれも閉領域をもつ。

表１に、学習の収束に要した中間ユニット数n_2と学習回数n_tを示す。これより、単調なアフィン関数（ア）と（ウ）に対しては、いずれの識別法も、たった１つの識別表面で関数の特徴を近似表現できることが分かる。(a)、(b)ともに、シグモイド関数の遷移領域中央付近のほぼ線形な入出力関係を利用して直線や平面を表現していた。（ウ）－(b)の場合には、中心が定義域の外部にある半径の大きな円の弧を用いて、定義域における直線を近似的に実現していた。(a)に比べ(b)の方が学習回数が多いが、これは、大きな値を解とする中心や半径の探索のため、変数空間の広い範囲を解探索しなければならないからだと考えられる。一方、閉領域をもつ（イ）と（エ）に対しては、予想どおり、(a)がより多くの中間ユニットを必要とし、入力次元が大きいほどその傾向が大きいことが確認できた。以上をまとめる。

テスト関数	(a)超平面識別	(b)超球面識別
(ア)	$n_2=1$ $n_t=1{,}900$	$n_2=1$ $n_t=26{,}000$
(イ)	$n_2=3$ $n_t=4{,}300$	$n_2=2$ $n_t=800$
(ウ)	$n_2=1$ $n_t=231{,}000$	$n_2=1$ $n_t=959{,}000$
(エ)	$n_2=8$ $n_t=294{,}200$	$n_2=1$ $n_t=511{,}900$

表１

■閉じた識別曲面の場合、閉領域の表現だけでなく、定義域において曲面の一部分だけを使うことにより、開領域の表現も可能である。その場合、変数空間の広い範囲を解探索する必要があるため、超平面の場合に比べて学習回数が多くなる傾向がある。

■超平面識別の場合、閉領域を表すには、領域の形・要求される精度・入力次元などに応じて、閉じた識別曲面の場合に比べ、かなり多くの中間ユニットが必要となる。特に、入力次元が大きい場合には、学習の収束は困難になると予想される。

■ＨＡ学習の収束のためには、変数の探索範囲を解を含む極力小さな領域に限定することにより無駄な探索を避けることが重要であり、そのためには、用いる演算方法や変数の役割に応じて適切な拘束条件を与えることが必要であった。

４．おわりに

ＮＮのＨＡ学習法において、超平面識別と超球面識別の２種類のユニット演算処理方法を用い、学習効率の違いを比較した。今後、今回の結果をふまえ、より適切なユニット演算処理方法と拘束条件の与え方を検討し、実用に耐え得る性能の実現を目指す。

参考文献

[1]新上，大田原："ニューラルネットワークのＨＡ学習の関数依存特性：偏差値関数の場合"、情報処理学会第 62 回全国大会発表、2-73（2001)．

[2]河野，下川："ニューラルネットワークのＨＡ学習の関数依存特性：高木関数"、情報処理学会第 62 回全国大会、2-87（2001)．