学習前にパターン認識用ネットの収束を判定する方法

7E−3

阿部重夫　鹿山昌宏　武長寛　北村忠明
(株) 日立製作所　日立研究所

1．はじめに

教師データを用いて逆伝播法で学習することにより、自動的にパターン判別ネットを構成できることから、パターン認識問題に多層ネットの適用が進められている。しかしながら多層ネットの学習に時間がかかり、また学習が収束しないとき、なぜ収束しないかを調べるためには、試行錯誤によらざるを得ないという問題がある。

このため、本論文では文献1), 2)で導いたニューラルネットの合成条件に基づいて、学習前にパターン認識用ネットが収束するか否かの判定を行う方式を提案し、数字認識でその妥当性を検証する。

2．３層ニューラルネットが構成できる条件

文献1), 2)によれば、もし全てのクラスが単一領域に分離されるとき、即ち、任意のクラスの入力データがいくつかの超平面の同じ側にあり、他のクラスのデータがこの領域にないとき、識別ネットは３層ニューラルネットで合成できる。学習の安定な収束と高い汎化能力を得るためにはこの条件は必要である。この条件はまた任意のクラスの入力データが凸超多面体の内部にあり、これらの超多面体が交差しないことと等価である。凸超多面体は、あらゆる組の入力データを全て線分でつなぎ、これらの線分で張られる凸超多面体の内部の線分を削除することにより得られる。しかしこの手順をそのまま実行すると大変なので、凸超多面体を入力変数の軸に平行な表面を持つ凹超多面体で近似することにする。そしてこの凹超多面体が交差しなければ、３層ニューラルネットが合成できると仮定する。この仮定を確実のものとするため、各クラス対の分離度をクラス間の入力データの最短距離で定義し、ある程度の分離度があるとき分離できるとする。

つぎにクラス i の超多面体の構成法について述べる。ここで m_k^i, M_k^i を入力変数 k に関するクラス i の入力データ最大値および最小値とする。そこで、クラス i にたいして、超多面体をつぎのように定義する。

$$P_i = \{x \mid m_k^i \le x_k \le M_k^i \text{ for } k = 1,\ldots,n\} \quad (1)$$

もし、超多面体 P_i, P_j が交差しないとき、即ち $P_i \cap P_j = \phi$ のときは、クラス i , j は分離可能である。このときは明かにどれかの入力変数で分離されていることになる。もし $P_i \cap P_j$ のなかにクラス i, j の入力データがないときあるいはどちらかのクラスの入力データがないときはそれらは分離可能である。交差領域の中に両方のクラスのデータが存在するときは、つぎのように交差領域を縮退することにより判定することにする。

(i) $m_k^j \le m_k^i \le M_k^j \le M_k^i$ のとき
交差領域の m_k^i を交差領域の入力変数 k に関するクラス i の最小値にする。

(ii) $m_k^i \le m_k^j \le M_k^i \le M_k^j$ のとき
交差領域の M_k^i を交差領域の入力変数 k に関するクラス i の最大値にする。

(iii) $m_k^j \le m_k^i \le M_k^i \le M_k^j$ のとき（図１参照）
交差領域の m_k^i, M_k^i を交差領域の入力変数 k に関するクラス i の最小値および最大値にする。

ついで新たに交差領域中に作られた超多面体が交差するかどうか調べる。もし交差しなければ、クラス i, j は分離可能である。もし交差するならば、上記の手順を交差しなくなるか、交差する領域が縮退しなくなるまで繰り返す。

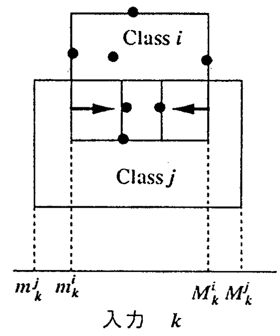

図１　交差領域の縮退

Determining the Convergence of Pattern Classification Neural Networks prior to Learning
Shigeo Abe, Masahiro Kayama, Hiroshi Takenaga and Tadaaki Kitamura
Hitachi, Ltd.

ここでクラスi, j が入力変数の集合 I = {1,..., m} に関して分離可能とは、上記の超多面体が交差しないことをいい、交差するときは分離不可能ということにする。またクラスi , j の入力変数の集合 I に関する分離度 $S_{ij}(I)$ を分離可能のときクラス間のデータの最小値とし、分離不可能のとき 0 とする。即ち

$$S_{ij}(I) = \min |x^i - x^j| \quad \text{クラス } i, j \text{ が分離可能}$$
$$= 0 \quad \text{クラス } i, j \text{ が分離不可能} \qquad (2)$$

ここでは x^i はクラス i の入力データに対応するn次元ベクトルで、最小値は二つのクラスの全てのデータの組合せについて取るものとする。また分離不可能なクラス i, j について、分離不可能度 $N_{ij}(I)$ を

$$N_{ij}(I) = \min (n_i, n_j) \qquad (3)$$

で定義する。但し n_i はクラス i のデータで交差領域にある数である。最小値を取った理由は、最小値を取るクラスの交差領域のデータをこの領域から移せば交差が解消するからである。

任意のクラスを分離可能とするために、分離余裕度 α_s (> 0) を導入し

$$S_{ij}(I) \geq \alpha_s \qquad (4)$$

を要請する。もしクラスi, j において、(4)が成立しないとき、このクラスは分離が難しいと判定する。もし$S_{ij}(I)$ が 0 なら$N_{ij}(I)$ によりどの程度難しいか判定できる。これらの判定を教師データについて行なえば、学習がしやすいか、難しいかの判定ができる。

上記手法の有利な点は、与えられた学習データと入力変数とで３層識別ネットが合成できるかどうか、学習の前に判定できることにある。更にまた分離しにくいクラスをも同定できることにある。

3 入力変数の選択

この節では学習の収束性を低下させることなく入力変数を削減する方法について述べる。それには、任意の$S_{ij}(I_s)$が、指定の余裕値より大きくなるような入力変数の集合I_sを求めることである。まず最初に$S_{ij}(I)$の最小値S_{min}を求める。S_{min}が０のとき、学習の収束を改善するために、適当な入力変数を加える必要がある。S_{min}が正とし、S_{min}より小さいα_sを設定する。この場合の入力の選定はまず集合 $I_s = I$としててI_sからある変数を削除して全てのクラスの組が分離余裕度 α_s 以上であればその変数を削除したままとし、そうでなければその変数をI_sに戻す。これを全てI_sの変数について行えばよい。

4 数字認識での評価

これまで述べた入力変数の選択方式の効果を 12の特徴量から数字認識を行なう数字認識システム[3]で評価した。

教師データとして 200データを選んだ。ニューラルネットは収束の判定を 0.1 として初期の重みを -0.1 から 0.1 の間に一様乱数で設定して、逆伝播法で学習した。初期値の影響を除くため、収束回数を 4000回として100 回学習して平均の認識率を求めた。このとき用いた計算機は61MIPSの計算機である。

12 入力のS_{min}は0.32であり、３層ネットで合成できると判定できる。そこで表１のようにα_sを変えて中間層ニューロン数を６にして３層ネットで学習した。なお、このときの中間層ニューロンの数は文献4)の方法で設定した。表により、α_sが小さくなると、削除される入力変数の数が増え、また学習に要する時間も増えている。$\alpha_s = 0.1, 0.05$のとき学習は収束していない。

表１　入力変数の選定と学習結果

選定			学習	
α_s	削除入力	cpu 時間 (秒)	平均認識率 (%)	平均cpu時間 (秒)
-	none	-	100	5.9
0.3	2,3,4,7	4.4	100	9.0
0.2	2,3,4,5	6.7	100	14.7
0.1	1,2,3,5,6,7,12	6.7	97.9	59.7
0.05	1,2,3,4,5,6,7,8	6.7	98.1	64.2

5．おわりに

パターン認識用ニューラルネットの構成理論に基づいてニューラルネットの学習の収束を事前に判定し、入力変数を選定する方式について述べ、その有効性を数字認識で確かめた。

参考文献

1) S. Abe, M. Kayama, and H. Takenaga, "How Neural Networks for Pattern Recognition Can Be Synthesized," Journal of Information Processing, Vol. 14., No. 3. 1991.

2) S. Abe, M. Kayama, and H. Takenaga, "Synthesizing Neural Networks for Pattern Recognition," Proc. IJCNN-91, Singapore, pp. 1105-1110, Nov. 1991.

3) 武長寛他、" 感度解析を用いたニューラルネットの入力層の最適化とその数字認識への適用" 、電学論文誌 D、平成３年１月、Vol. 111-D, No.1, pp. 36-44, Jan. 1991.

4) M. Kayama, S. Abe, H. Takenaga, Y. Morooka, "Constructing Optimal Neural Networks by Linear Regression Analysis," Proc. Neuro-Nimes '90, pp. 363-376, November 1990.