4K－5

階層形ニューラルネットワークにおける学習データ選択法*

原 一之 中山謙二
金沢大学大学院自然科学研究科 金沢大学工学部電気・情報工学科
E-mail：nakayama@ec.t.kanazawa-u.ac.jp

1 はじめに

パターン分類問題に誤差逆伝播法 (BP)[1] を用いた場合、適切な学習データと、適切なネットワークを用いて学習することにより、学習に用いたパターンに共通する特徴や規則を自動的に抽出することが期待できる。このように、BP 法は分類規則を与えることなしに、例題に基づいた分類を行なうことが出来る。このとき問題となるのは、適切な学習データの選択、つまり、学習データのサンプルの仕方と、サンプルの数、そして最適なネットワークの選択、すなわち各層のユニットの数と活性化関数の選択である。本稿では、学習データの選択法について提案し、さらに学習のある段階で提案方法によって学習データを減らすことにより、学習に要する計算量を減らすことが出来ることを示す。

2 分類とグループ境界

本稿ではデータの分類に階層形ニューラルネットワーク (MLNN) を用る。隠れユニットの活性化関数はシグモイド関数とする。学習アルゴリズムとしては誤差逆伝播 (BP) 法を用いる。

MLNN において、活性化関数 $\varphi(\cdot)$ とした場合、例えば 2 グループからなるデータデータ $X = \{X_1, X_2\}$ は、次の条件を満足したとき、 $\varphi(x) = [\varphi_1(x), \varphi_2(x), \ldots, \varphi_M(x)]^T$ で分離可能である。

$$\left.\begin{array}{ll} w^T\varphi(x) & > 0, x \in X_1 \\ w^T\varphi(x) & < 0, x \in X_2 \end{array}\right\} \quad (1)$$

ここで、超平面 $w^T\varphi(x) = 0$ は φ 空間における分離面を表す。2 層のネットワークの場合、x は隠れ層入力と考えられる。ベクトル wは隠れ層・出力層間の結合である。ベクトル wは学習によって得られ, 真のベクトル wの推定値$\hat{w}$となる。このように、MLNN 内部では隠れ層の出力空間上でデータを分離しており、分離面を挟むように、境界面が各グループに対して 1 枚づつ形成される。

3 境界付近のデータによる学習

本稿で用いた MLNN は、2 層の MLNN で、隠れ層、出力層の活性化関数は [-1,1] および [0,1] のシグモイド関数である。2 章によれば、分類は出力ユニットの入力で完結しているが、飽和関数を用いることにより、収束を早め、かつ学習を安定させる効果が期待できる。

一般に BP 法を用いた学習は、最急降下法により誤差が小さくなるよう、結合係数を修正するが、誤差が小さくなると最急降下の傾きが小さくなるため、誤差の収束が遅くなる。この時、データ全体に対する誤差はある程度小さくなっており、少数のデータの誤差だけが大きいと考えられる。したがって、初期学習の後、誤差減少に寄与するデータのみを選択し、追加学習することによって、学習に要する計算量を削減できると考えられる。以下、その方法について説明する。

3.1 誤差の分担

BP を用いた分類問題では、入力データに対する所望出力と出力層出力の差を誤差とする。そして、出力層における誤差に比例して、ユニット間の結合係数を修正する。分類問題においては、分類数だけの出力ユニットを用意し、所望出力として、入力データのグループに対応する出力だけが "1"、他のユニットは "0" を割り当てる。データに対する誤差の分布は学習の進度、活性化関数により異なる。学習の初期においては、全てのデータが同程度の誤差値であるが、学習が成功する場合、学習が進むにつれて大多数のデータに対しては誤差は十分小さくなり、少数のデータの誤差が、全パターンの誤差を担うようになる。各グループの全体のサンプル数を N、誤差が小さくなったサンプルの数を K とすると、誤差の大きなサンプル数 M は $M = N - K$ で与えられる。グループ全体のパターン p に対する誤差を E_{Tp} とすると

$$\sum_{p=1}^{N} E_{Tp} = (\sum_{p=1}^{K} E_{Sp} + \sum_{p=K+1}^{N} E_{Lp}) \quad (2)$$

ここで、E_{Sp}、E_{Lp} は p 番目のパターンに対する誤差の小さいデータの誤差、誤差の大きいデータの誤差である。結合係数の修正は、E_Tを用いて計算されるので、小さい誤差の総和が大きい誤差の総和に比較して十分小さい時は、結合係数の修正量はほぼ $\sum_{p=k+1}^{N} E_{Lp}$ で決定されると考えることができる。

誤差の小さくなったデータに対しては、結合件数の修正がほとんど行われないことから、これらのデータを用いなくても、全てのデータを用いて学習した場合と同様の収束状態に導かれることが分かる。

*Training Data Selection for Multilayer Neural Netowrk, Kazuyuki HARA Kenji NAKAYAMA, Graduate School of Nat. Sci. & Tech., Faculty of Engineering, Kanazawa University, 2-40-20 Kodatsuno Kanazawa 920

3.2 学習誤差と境界

出力層入力は、隠れ層出力空間における分離面との内積で、入力 “0” は分離面である。今、出力ユニットの所望出力を “1” とした場合、入力の正の領域が望ましい領域である。この時、誤差の小さいデータは境界面から離れた領域に分布し、比較的誤差の大きなデータは “0” 近傍または負の領域にある。

また、分離面は学習が進むにしたがって、誤差の大きいデータがある方へ移動する。このとき、3.1 節で述べたように、誤差の大きいデータが結合係数の修正、つまり、分離面の構成に対して支配的になる。よって、この点からも誤差の小さいデータは学習への寄与が小さい。以上より、誤差の小さいデータ、つまり境界面から離れたデータを削除することにより、同じ epoch 数の学習で学習が収束するとした場合、学習データが削減された分、学習に要する計算量は削減できることがわかる。

3.3 問題点

出力ユニット入力の分布は学習が進むにしたがって wが絶対値的に大きくなるため、飽和領域に近づく。したがって、学習時の誤差が小さくなる前に初期学習を止め、データを選択した場合、多くのデータが残りデータ選択の効果が減る。一方、誤差が収束してしまってからでは、すでに多くの学習を行なっているので、やはり効果が低い。よって、学習を中断する基準となる誤差の選定が重要である。

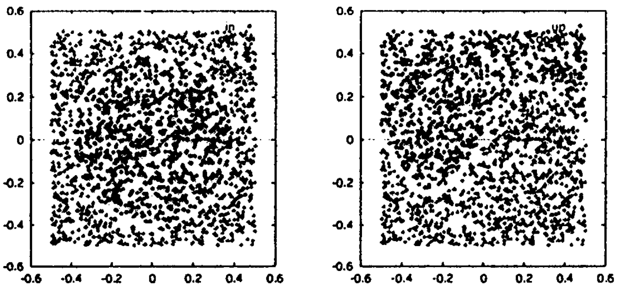
また、再学習の収束性を高めるためには、境界を埋め尽くすようにデータが残ることが望ましい。しかし、学習するデータの分布が異なる場合、境界構成の難しさに差が生じる。したがって、1 グループの境界に対応する他グループのデータが常に選択される保証はない。したがって、本稿ではデータ選択によって残ったデータに対し、次式のようなペアリングの操作を行ったない、選択データに追加する。

if $x \in X_1$ then add x'= $min\{||x - x'||\}, x' \in X_2$
if $x \in X_2$ then add x'= $min\{||x - x'||\}, x' \in X_1$

4 シミュレーション

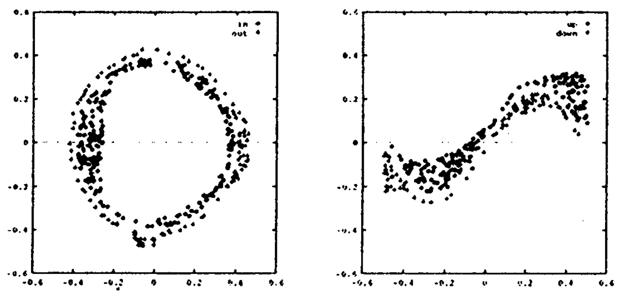
図 1に 2 つの問題を示す。図 1(a) を問題 1、(b) を問題 2 である。分類数は 2、データ数は各グループ 1000 個用いた。ネットワークは入力ユニット 2、隠れユニット 6、出力ユニット 2 とした。この例では、問題 1 は 1 つのグループは閉領域、他方は開領域、問題 2 は 2 グループとも開領域となっている。学習の難しい場合、誤差は大きいので、境界近傍のデータが多くなり、ある出力層入力値 th 以下となるデータが増える。

初期学習で、MSE(Mean Square Error)<0.05 まで学習した後のネットワークを用い、$th = 1.0$ とした時、問題 1 では円内が 252 個, 円外は 78 個残り、問題 2 では境界の正弦波の上の領域に 191 個、下の領域に 80 個残った。これより、問題 1 は学習の難しさに差があると考えられる。選択の後、ペアリングを行った結果を図 2に示す。今回、MSE および th は経験的に決定した。問題 1 において、途中で止めずに $MSE < 0.001$ まで学習した場合と提案法を比較したところ、途中で止めずに学習した場合、1980 epoch、提案法では 2500 epoch 要した。MSE=0.001 は所望出力 “1” に対して “0.968” を出力することに相当する。提案法で選択されたデータは、ペアリングを行なった後 662 個残るので、データ数 × epoch を計算量とすると、約 53%削減された。問題 2 でも同様の条件にて 60%計算量が削減できた。



(a) circle in square　　(b) sinusoidal in square

図 1: 学習に用いた問題



(a) circle in square　　(b) sinusoidal in square

図 2: 境界近傍に残ったデータ

表 1: シミュレーション結果: 問題 1

学習法	分類率 1	分類率 2	MSE
選択無し	−	100	0.000989
選択あり	94.7	100	0.000995

分類率 1:初期学習
分類率 2:追加学習

5 まとめ

階層形ニューラルネットワークを用いた分類問題において、出力層の入力の値により、学習に有効なデータを選択する方法を提案した。今後は最適な境界を検出するための学習停止基準を決定する方法、本提案方法を学習するデータの領域が変化する場合の追加学習への適用について検討する予定である。

参考文献

[1] D.E.Rumelhart and J.L.McCelland et. al., Parallel Distributed Processing, MIT Press, 1986

[2] 福水, 渡辺, “関数近似問題における最適学習設計と予測誤差”, 信学技法,NC94-98, pp.173-180, 1995.3.