

額賀良平†

天方貴久‡

原一之†

† 東京都立工業高等専門学校 電子情報工学科

‡ 東京農工大学工学部情報コミュニケーション工学科

## 1 はじめに

階層型ニューラルネットワーク (以下 階層型 NN) とは学習によってパラメータである結合係数を修正し、期待する入出力関係を得る学習機械である。学習に必要な学習データは入力  $\mathbf{x}_p$  とそれに対する期待出力  $t_p$  との組からなる。また、次式によって二乗平均誤差 (以下 MSE) が定義されている。

$$E = \frac{1}{n} \sum_p \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2 = \frac{1}{n} \sum_p e_p \quad (1)$$

ここで  $o_{pj}$  はネットワークの出力である。

## 2 提案法 1

階層型 NN において、誤差逆伝播法 (以下 BP 法) によってバッチ型学習を行ったとき、学習が進むにつれ MSE の大きなデータは少なくなり、MSE の小さなデータが支配的となる。BP 法において、出力層における結合係数の修正量は以下の式で求められる。

$$\Delta_p \omega_{ij} = \eta (t_{pj} - o_{pj}) f'_o(\text{net}_{pj}) o_{pi} \quad (2)$$

式 (1)(2) より、MSE が大きいほど結合係数の修正量は大きいことがわかる。しかし、バッチ型学習での結合係数の修正量は、各データの修正量を平均した値によって決まるので、MSE の小さなデータがに対し修正が行われ、MSE の大きなデータに対して行われない。

そこで我々は、MSE の大きなデータと MSE の小さなデータの数的バランスをとるような学習法を考案した (以下 提案法 1)。しかし、MSE の小さなデータを全く学習に用いないと学習が困難になるので、結合係数の修正に用いるデータを、MSE の大きさに応じて確率的に選ぶようにした。このとき以下の確率密度関数を用いる。

$$\text{Prob}(e_p) = \frac{\exp(m \cdot e_p) - 1}{\exp(m) - 1} \quad (3)$$

以下、この確率密度関数を修正関数と呼ぶことにする。ただし、 $e_p$  は  $p$  番目の学習データの MSE を表す。修正関数は図 1 の様な関数となる。

\*A learning by stochastic update of connection weights, Nukaga Ryouhei, Yoshihisa Amakata, Kazuyuki Hara, †Tokyo Metropolitan College of Technology, 1-10-40 Higashi-oi, Shinagawa, Tokyo 140-0011 Japan. ‡Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588

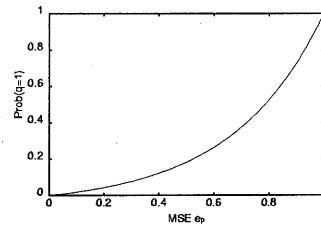


図 1: 修正関数

提案法 1 で学習を行っていくと MSE の大きなデータに対して重点的に学習が進むので、最大値  $\arg \max\{e_p, p = 1 \dots P\}$  が小さくなり、結合係数の修正に用いられるデータ数が低下してしまう。よって修正関数の  $e_p$  は、最大値でスケールされた値を用いる。

### 2.1 計算機シミュレーション

通常の BP 法と提案法 1 で次のようなシミュレーションを行った。学習データは図 2 のように円の内側と外側の 2 クラスに分類するようなデータを用意する。ただし円は、円の外側の面積対内側の面積を 3 対 1 に分割するような半径に設定した。また、各クラスの領域のデータの分布を変え、図 3~図 5 のような 3 通りの学習データに対し学習を行った。このときのネットワークは入力層、中間層 1 層、出力層の 3 層で構成し、それぞれのユニットは 2、4、2 個とした。このとき使用した修正関数では  $m = 3$  とした。シミュレーションは全データの分類率が 100% に達するか、学習回数が学習限度回数に達するまで学習を繰り返した。初期結合係数を乱数によって決定しているため、乱数種を変えて 10 回シミュレーションを行い計算量、分類率の平均値を計算した。

### 2.2 結果

結果は次のようになった。

	BP 法		提案法 1	
	分類率	計算量	分類率	計算量
一様分布のデータ	0.9946	20000.0	0.9926	5116.4
ガウス分布のデータ	0.9994	19835.7	1.0000	3841.8
数に偏りのあるデータ	0.9341	100000.0	1.0000	8169.9

計算量は全学習データを用いて結合係数を 1 回修正し

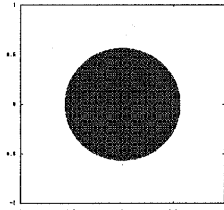


図 2: 学習データの概略分布

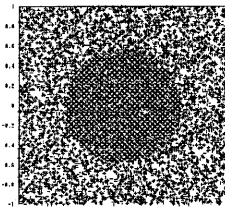


図 3: 一様分布のデータ

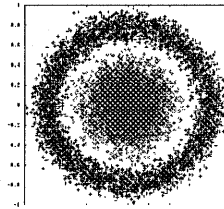


図 4: ガウス分布のデータ

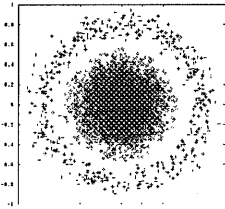


図 5: 数に偏りのあるデータ

たときの計算量を 1 とした。一様分布のデータを学習させたとき、提案法 1 は通常の BP 法と同等の分類性能を持ちながらおよそ 1/4 計算量で学習を行うことができた。ガウス分布のデータに対しても同様に少ない計算量で同等の分類性能が得られたといえる。また、数に偏りのあるデータにおいて BP 法の分類率は 93.41% であったが、これは数の少ない方のクラスをほとんど分類できなかったことを意味する。これに対し、提案法 1 は 1/2 以下の計算量で高い分類性能を発揮した。

これらのことから提案法 1 は少ない計算量で BP 法と同等の分類性能を持ち、数に偏りのあるデータに対しては BP 法以上の分類性能を持つことがわかった。

### 3 提案法 2

提案法 1 によって学習を行った場合、通常の BP 法と比較すると大幅に計算量を削減することができた。よって、さらなる計算量削減のために全学習データの中から学習に使用するデータをサンプリングし、サンプリングされたデータのみで学習を行うような学習法を考案した(以下 提案法 2)。

提案法は、学習前に全学習データの中からいくつかの学習データをサンプリングし、データの誤差分布の推定を行う。サンプリングされたデータを用いた学習での結合係数の修正量は全学習データを用いた場合とは異なる。しかし、サンプリングの数が減少してもデータの統計的性質が同等であれば、全学習データを用いた学習での修正量に近い値で結合係数の修正が行われることが期待できる。

サンプリングの方法としては、図 6 の様に、全学習データの中から学習毎にある一定の割合だけデータをサンプリングする方法をとった。この方法でサンプリングを行えばサンプリングした学習データから情報を有効に

利用して、全学習データを用いて学習した場合と同等の結合係数の修正が行われる事が期待できる。

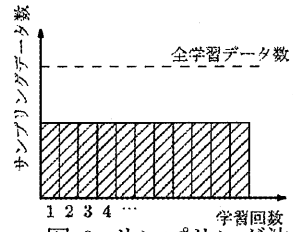


図 6: サンプリング法

### 3.1 計算機シミュレーション

提案法 2 において、2.1 と同様のシミュレーションを行った。今回のシミュレーションでは全データの 44.6% をサンプリングした。サンプリングしたデータは 10 回の学習毎にサンプリングし直した。

### 3.2 結果

シミュレーションの結果は次のようになった。

	提案法 2	
	分類率	計算量
一様分布のデータ	0.9909	2168.2
ガウス分布のデータ	0.9991	1723.0
数に偏りのあるデータ	0.9091	14581.0

一様分布のデータとガウス分布のデータに対しては、提案法 2 は BP 法と同等の分類性能を保ったまま提案法 1 よりも少ない計算量で学習を行えた。しかし数に偏りのあるデータに対しては、全く分類することができなかった。

これらのことから、提案法 2 は一様分布やガウス分布のデータのように分布に偏りのあるデータに対しては BP 法と同等の分類率を持ちながら、提案法 1 よりも少ない計算量で学習が行えるが、数に偏りのあるデータに対しては学習が困難であることがわかった。

### 4 今後の課題

提案法 2 は、サンプリングを行うことで数の少ないクラスのデータをさらに減らしていたために、分類性能が悪かったものと予想される。今後の課題として、データ数の少ないクラスからはデータを減らさないようにして学習させたい。

### 参考文献

- [1] C.Cachin, *Pedagogical pattern selection strategies Neural Networks*, vol. 7, no. 1, pp. 175-181, 1994.
- [2] 天方, 原, "誤差分布を考慮した動的学習法", 信学ソ大(情報・システム), pp. 14, 1999.
- [3] 額賀, 天方, 原, "誤差分布を考慮した動的学習法における分布の推定", 進学東京支部学生会論文集, pp. 23, 2000.