

環境の変化に適応できる複製・競合動径基底関数ネットワーク

奥原 浩之[†]佐々木浩二^{††}尾崎 俊治^{††}

Reproductive and Competitive Radial Basis Function Networks Adapting to the Dynamical Environments

Koji OKUHARA[†], Koji SASAKI^{††}, and Shunji OSAKI^{††}

あらまし 関数近似問題やパターン識別に適したニューラルネットワークの一つに動径基底関数ネットワークを考えることができる。動径基底関数ネットワークは階層型ニューラルネットワークに比較してニューロンごとの局所的な学習が可能であるなどの優れた点をもつ。しかし、動径基底関数ネットワークでは未知の非線形関数を近似するため、あらかじめ必要なニューロン数が不明である。このことが学習の遅延化や過学習の問題を引き起こしている。これらの問題を解決するために、我々は先に競合動径基底関数ネットワークを提案した。競合動径基底関数ネットワークでは、シナプス結合荷重間の競合を考慮したシナプス可塑性方程式を利用することにより学習の効率化を図ることができた。ところが、競合動径基底関数ネットワークは不必要的動径基底関数を削除する能力をもつものの、必要とされる動径基底関数を追加する能力は備えていなかった。そこで本論文では、シナプス可塑性方程式に関する考察から、必要な動径基底関数を効率的に追加する手法を提案する。この手法では、学習が収束した時点で得られている動径基底関数の特性の一部を新たに追加される動径基底関数が引き継いでいる。そこで、本手法を競合動径基底関数ネットワークに組み合わせたニューラルネットワークを複製・競合動径基底関数ネットワークとして提案する。シミュレーションでは、複製・競合動径基底関数ネットワークを関数近似に適用し、累積2乗誤差関数を効率良く、より小さくできることを示す。また、従来の動径基底関数ネットワークに比較して高速に学習できることも示す。

キーワード シナプス可塑性方程式、競合動径基底関数ネットワーク、自由エネルギー、複製

1. まえがき

ニューラルネットワーク (Neural Networks : 以下, N.N.) の一つに動径基底関数ネットワーク (Radial Basis Function Network : 以下, RBFN) [1] がある。RBFN はニューロンごとの局所的な学習が可能であり、ほかの N.N. に比較して学習が高速であることが知られている。また、十分な数のニューロンを用いた RBFN は任意の非線形関数を有界閉集合上で一様に近似できることができることが関数近似定理 [2] で示されている。

一般に、未知の非線形関数を近似するために必要なニューロンの数があらかじめには不明である。もし、N.N. に冗長なニューロンが多数存在する場合は、学

習の遅延や過学習の問題を生じることとなる。そこで、冗長なニューロンを削除する手法が提案されている [3], [4]。これに対し、N.N. に関数近似に必要な数のニューロンが存在しない場合は、関数近似をすること自体が不可能となる。そこで、新たに必要なニューロンを追加する手法が提案されている [5]。これら従来の研究には、しきい値などを考え動径基底関数の削除と追加を行うものもある。ところが、このような手法では基準となるしきい値の決定自体が困難であることが予想できる。また、教師信号が動的に変化する環境では削除する手法と追加する手法を組み合わせて学習を行わなければならない。当然それぞれにとって良い手法を、ただそのまま組み合わせただけでは、動径基底関数の数が振動するなどして望ましい結果が得られるとは限らない。

我々は先に、冗長なニューロンを削除できるシナプス可塑性方程式を導出し、これを適者生存型学習則としてシナプス結合荷重の更新則に適用した競合動

[†] 広島県立大学経営学部経営情報学科、庄原市

Department of Management and Information Sciences,
Hiroshima Prefecture University, Shyobara-shi, 727-0023
Japan

^{††} 広島大学工学部第二類(電気系), 東広島市

Faculty of Engineering, Hiroshima University, Higashi-Hiroshima-shi, 739-8527 Japan

径基底関数ネットワーク (Competitive Radial Basis Function Network: 以下, CRBFN) [6] を提案した。CRBFN の特長は望ましい出力と相關が高い入力を伝達しているシナプス結合荷重が生き残り、生き残ったシナプス結合荷重に関係する入力ニューロンの入出力特性のみが調節されることである。そのため、CRBFN では競合により冗長なニューロンを消滅させることができ、その結果、学習の高速化と過学習の回避が行われる。

しかしながら、CRBFN でも教師信号が変化するような環境の変化には対応しきれていなかった。その理由は CRBFN には新しい動径基底関数を追加する能力がないからである。ここでいう環境の変化とは入出力間の写像を与える関数そのものが変化する場合や、既に観測され学習に用いられていた入出力の組が不要となり取り除かれたり、新たに観測された入出力の組が学習に用いられたりするような変化などを想定している。そこで本研究では、まず新しい動径基底関数を追加する手法を提案する。この手法はシナプス可塑性方程式 (Delta ルール、適者生存型学習則) に関する考察から得られるものであり、必要な動径基底関数を効率的に追加することができる。そして、我々が先に提案した CRBFN にこの手法を組み合わせたニューラルネットとして複製・競合動径基底関数ネットワーク (Reproductive CRBFN: 以下, RC-RBFN) を提案する。この RC-RBFN は、環境の変化に適応する能力を備えたものとなっている。本論文では、最終的に得られる結果の有効性から適者生存型学習則に対する動径基底関数の複製アルゴリズムについて議論を展開しているが、同様な議論が従来の Delta ルールに対しても可能であることも付録で述べる。

本論文の構成は次のとおりである。2. では CRBFN の概要について述べる。3. ではシナプス可塑性方程式に関する考察から、動径基底関数のパラメータの従う確率密度関数の導出を行う。4. では導出された確率密度関数を利用した動径基底関数の複製アルゴリズムを提案する。5. では提案した RC-RBFN の性能をシミュレーションにより確認する。6. ではまとめと今後の課題について述べる。

2. 競合動径基底関数ネットワークの概要

RBFN は非線形関数 $\eta(\mathbf{x})$ を動径基底関数の足し合いで近似するニューラルネットワークである。動径基底関数としては規格化されたガウス型活性関数などが用い

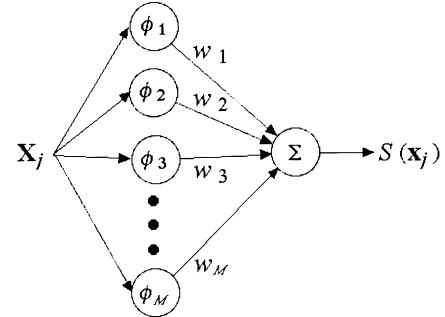


図 1 RBFN の構造

Fig. 1 The structure of the RBFN.

られる。 M 個の入力ニューロンと 1 個の出力ニューロンからなる RBFN は図 1 のような構造をもつ。 d 次元の第 i 入力ベクトル $\mathbf{x}_i \in R^d$, ($i = 1, 2, \dots, N$) はすべての入力ニューロンに入力される。第 j 入力ニューロン ($j = 1, 2, \dots, M$) はパラメータ ϕ_j をもつ。パラメータ ϕ_j は平均ベクトルと共に分散行列の集合 $\{\mathbf{m}_j, \Sigma_j\}$ であるものとする。ここで, $\mathbf{m}_j = [m_j^1, m_j^2, \dots, m_j^d]^T$ であり, Σ_j はその逆行列 Σ_j^{-1} の第 kl 要素に σ_j^{kl} をもつ $d \times d$ の行列である。また, Σ_j は正定値対称行列である。第 j 入力ニューロンは入力ベクトル \mathbf{x}_i に対して

$$\xi(\mathbf{x}_i, \phi_j) = \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \mathbf{m}_j)^T \Sigma_j^{-1} (\mathbf{x}_i - \mathbf{m}_j) \right\} \quad (1)$$

を出力する。ここで、添字の T はベクトルの転置を示す。以後、このような出力を有する入力ニューロンのことを動径基底関数ということとする。出力値 $\xi(\mathbf{x}_i, \phi_j)$ はシナプス結合荷重 w_j を通して出力ニューロンへ伝達され、出力ニューロンでこれらは足し合わされ

$$s(\mathbf{x}_i, \mathbf{w}, \phi) = \sum_{j=1}^M w_j \xi(\mathbf{x}_i, \phi_j) \quad (2)$$

が outputされる。ここで, $\mathbf{w} = [w_1, w_2, \dots, w_M]^T \in R^M$ であり, ϕ で集合 $\{\phi_1, \phi_2, \dots, \phi_M\}$ を表す。ニューラルネットワークによる関数近似は、非線形関数 $\eta(\mathbf{x})$ をネットワークの出力 $s(\mathbf{x}, \mathbf{w}, \phi)$ で表すことである。そのため、RBFN による関数近似は累積 2 乗誤差関数

$$E(\mathbf{w}, \phi) = \frac{1}{2} \sum_{i=1}^N E(\mathbf{x}_i, \mathbf{w}, \phi) \quad (3)$$

の値を減少させることにより実現される。ここで、

$$E(\mathbf{x}_i, \mathbf{w}, \phi) = \{\eta(\mathbf{x}_i) - s(\mathbf{x}_i, \mathbf{w}, \phi)\}^2 \quad (4)$$

は 2 乗誤差関数である。つまり、RBFN が学習により獲得しなければならないのは、第 j 動径基底関数のシナップス結合荷重 w_j 、パラメータ \mathbf{m}_j 並びにパラメータ Σ_j である。

ここで、従来の RBFN と CRBFN の学習アルゴリズムの相違について述べる。一般的 RBFN の学習アルゴリズムは式(3)の累積 2 乗誤差関数に Delta ルール [7] を適用した

$$\Delta w_j = -\epsilon \frac{\partial E(\mathbf{w}, \phi)}{\partial w_j}, \quad (5)$$

$$\Delta m_j^k = -\epsilon \frac{\partial E(\mathbf{w}, \phi)}{\partial m_j^k}, \quad (6)$$

$$\Delta \sigma_j^{kl} = -\epsilon \frac{\partial E(\mathbf{w}, \phi)}{\partial \sigma_j^{kl}} \quad (7)$$

で与えられる。ただし、 ϵ は適当な正の定数であり、 m_j^k はパラメータ \mathbf{m}_j の第 k 要素である。また、 $\Delta w_j \equiv dw_j/dt$ 、 $\Delta m_j^k \equiv dm_j^k/dt$ 、 $\Delta \sigma_j^{kl} \equiv d\sigma_j^{kl}/dt$ である。ところで、未知の非線形関数を近似するために必要な動径基底関数の個数をあらかじめ知ることはできない。そのため一般に、RBFN では初期状態においていくつかの冗長な入力ニューロンを備えている。このことは、学習の遅延化や過学習を招く原因の一つであった。

CRBFN では、パラメータ \mathbf{m}_j 並びにパラメータ Σ_j の学習アルゴリズムは従来の RBFN と同じであり、式(6)、(7)により与えられる。しかし、シナップス結合荷重 w_j に対しては Dale 則を考慮したシナップス可塑性方程式である適者生存型学習則

$$\Delta w_j = \epsilon \left(\alpha_j(\phi) - \sum_{k=1}^M \gamma_{jk}(\phi) \mu_k w_k \right) w_j \quad (8)$$

が適用される[8]。シナップス結合荷重 w_j は Dale 則により、第 j ニューロンが興奮性であるなら正、抑制性であるなら負の値をとる。また、 μ_k は第 k ニューロンが興奮性 ($w_k > 0$) であるなら 1、抑制性 ($w_k < 0$) であるなら -1 となる Dale 則を考慮した識別子である。ここで、 $\alpha_j(\phi)$ は内的自然増加率であり

$$\alpha_j(\phi) = \sum_{i=1}^N \eta(\mathbf{x}_i) \xi(\mathbf{x}_i, \phi_j) \quad (9)$$

で定義される。 $\gamma_{jk}(\phi)$ は競争係数であり、第 j ニューロンと第 k ニューロンとの競合の効果を表し

$$\gamma_{jk}(\phi) = \sum_{i=1}^N \xi(\mathbf{x}_i, \phi_j) \xi(\mathbf{x}_i, \phi_k) \quad (10)$$

で定義される。また、CRBFN の出力は

$$s(\mathbf{x}_i, \mathbf{w}, \phi) = \sum_{j=1}^M \mu_j w_j \xi(\mathbf{x}_i, \phi_j) \quad (11)$$

で得られることとする。このとき、累積 2 乗誤差関数 $E(\mathbf{w}, \phi)$ は式(8)に対する Lyapunov 関数であることが

$$\begin{aligned} & \frac{dE(\mathbf{w}, \phi)}{dt} \\ &= \sum_{j=1}^M \frac{\partial E(\mathbf{w}, \phi)}{\partial w_j} \frac{dw_j}{dt} \\ &= -\sum_{j=1}^M \mu_j \left\{ \sum_{i=1}^N \eta(\mathbf{x}_i) \xi(\mathbf{x}_i, \phi_j) \right. \\ &\quad \left. - \sum_{h=1}^M \mu_h w_h \sum_{i=1}^N \xi(\mathbf{x}_i, \phi_j) \xi(\mathbf{x}_i, \phi_h) \right\} \frac{dw_j}{dt} \\ &= -\epsilon \sum_{j=1}^M \mu_j w_j \left(\alpha_j(\phi) - \sum_{h=1}^M \gamma_{jh}(\phi) \mu_h w_h \right)^2 \\ &\leq 0 \end{aligned} \quad (12)$$

により示される。

CRBFN の学習則は累積 2 乗誤差関数 $E(\mathbf{w}, \phi)$ の値を減少させるために第 j 動径基底関数のシナップス結合荷重 w_j を式(8)で更新する。学習中に $w_j \approx 0$ となった第 j シナップス結合荷重は消滅したものとして、生き残っているシナップス結合荷重、並びにそれらにより伝達される入力が変化を受ける動径基底関数のパラメータについてのみ学習を続ける。この学習則はシナップス結合荷重の更新則に特徴があるものの、平均ベクトルと共に分散行列の更新則は従来の最急降下法を用いている。

そこで、本研究では更に CRBFN において平均ベクトルの更新則を改良することにより、動径基底関数を複製する競合動径基底関数ネットワークを新たに提案する。

3. パラメータが従う確率密度関数の導出

ここでは、CRBFN の平均ベクトル、共分散行列とシナップス結合荷重が学習終了時にとる同時確率密度 $p(\mathbf{w}, \phi)$ を導出する。まず、シナップス結合荷重 w_j を

$$y_j^2 = \mu_j w_j \quad (13)$$

と変数変換する。 y_j の定義域は任意の実数である。このとき、式(8)は

$$\frac{dy_j}{dt} = \left(\frac{\alpha_j(\phi)}{2} - \sum_{k=1}^M \frac{\gamma_{jk}(\phi)}{2} y_k^2 \right) y_j \quad (14)$$

となる。ただし、 ϵ は省略した。式(14)は積分条件

$$\frac{\partial}{\partial y_k} \frac{dy_j}{dt} = \frac{\partial}{\partial y_j} \frac{dy_k}{dt} \quad (15)$$

を満たすためポテンシャル

$$V'(\mathbf{y}, \phi) = - \sum_{j=1}^M \int_a^{y_j} \left(\frac{\alpha_j(\phi)}{2} - \sum_{k=1}^M \frac{\gamma_{jk}(\phi)}{2} y_k'^2 \right) y'_j dy'_j \quad (16)$$

を考えることができ、変数 y_j の時間変化はポテンシャル $V'(\mathbf{y}, \phi)$ から、

$$\frac{dy_j}{dt} = - \frac{\partial V'(\mathbf{y}, \phi)}{\partial y_j} \quad (17)$$

で導くことができる。関係式(13)からポテンシャル $V'(\mathbf{y}, \phi)$ は

$$V(\mathbf{w}, \phi) = - \sum_{j=1}^M \left\{ \frac{\alpha_j(\phi)}{4} \mu_j w_j - \sum_{k \neq j}^M \frac{\gamma_{jk}(\phi)}{4} \mu_j w_j \mu_k w_k - \frac{\gamma_{jj}(\phi)}{8} w_j^2 \right\} \quad (18)$$

と書き直すことができる。この結果、

$$E(\mathbf{w}, \phi) = 4V(\mathbf{w}, \phi) + \frac{1}{2} \sum_{i=1}^N \eta^2(\mathbf{x}_i) \quad (19)$$

であることが示されるので、累積2乗誤差関数 $E(\mathbf{w}, \phi)$ の最小化はポテンシャル $V(\mathbf{w}, \phi)$ の最小化と等価であることがわかる。

今、式(14)に従う y_j はポテンシャル $V'(\mathbf{y}, \phi)$ の最急降下方向に更新される。その結果、ひとたび極小解に収束すると、そこから逃れることができなくなる。そこで、極小解から脱出させるための手法として、 y_j

の更新則を

$$y_j(t + \Delta t) = y_j(t) - \frac{\partial V(\mathbf{y}, \phi)}{\partial y_j} \Delta t + \sqrt{Q \Delta t} n_j(t) \quad (20)$$

のようにノイズを考慮し離散近似した見本過程で与えることが考えられる。ただし、 $n_j(t)$ は独立な確率変数であり、平均 0、分散 1 の正規分布 $N(0, 1)$ に従う。 Q は任意の正の定数である。このとき、学習終了時に CRBFN の平均ベクトル、共分散行列とシナプス結合荷重が満たす同時確率密度 $p_\beta(\mathbf{w}, \phi)$ は

$$p_\beta(\mathbf{w}, \phi) = Z_\beta^{-1} \exp\{-\beta V(\mathbf{w}, \phi)\} \quad (21)$$

で得ることができる[9]。ここで、 $\beta = 2/Q$ である。 Z_β は分配関数であり

$$Z_\beta = \int_{\mathbf{w}} \int_{\phi} \exp\{-\beta V(\mathbf{w}, \phi)\} d\mathbf{w} d\phi \quad (22)$$

で定義される。また、式(21)はポテンシャル $V(\mathbf{w}, \phi)$ と累積2乗誤差関数 $E(\mathbf{w}, \phi)$ の関係式(19)より

$$p_{\beta'}(\mathbf{w}, \phi) = Z_{\beta'}^{-1} \exp\{-\beta' E(\mathbf{w}, \phi)\} \quad (23)$$

と書き直すことができる。ここで、 $\beta' = (2Q)^{-1}$ である。また、 $Z_{\beta'}$ は分配関数である。シナプス可塑性方程式として Delta ルールを用いている従来の RBFN に關しても、同様にしてパラメータが従う確率密度関数が導出できることを付録で示す。

以上のようにして、パラメータが従う確率密度関数が導出できたことにより、与えられた条件のもとで累積2乗誤差関数 $E(\mathbf{w}, \phi)$ を最小とするパラメータの値が検出できることを示す。ここでは、教師信号 $\eta(x)$ を

$$\eta(x) = 3N(-1.5, 1) + 2N(1, 0.5) \quad (24)$$

で与えることとする。 $N(m, \Sigma)$ は平均 m 、分散 Σ の正規密度関数を表す。この教師信号を動径基底関数を一つ(シナプス結合荷重 $w = 1$ 、パラメータ $\Sigma = 0.2$)だけ用いて近似することを考える。この場合、近似しようとしている非線形関数 $\eta(x)$ の複雑さに対し、必要とされる動径基底関数が十分に存在していないため、累積2乗誤差関数 $E(w, \phi)$ を 0 にすること自体が不可能である。しかし、この動径基底関数のパラメータ m が従う条件付き確率密度関数

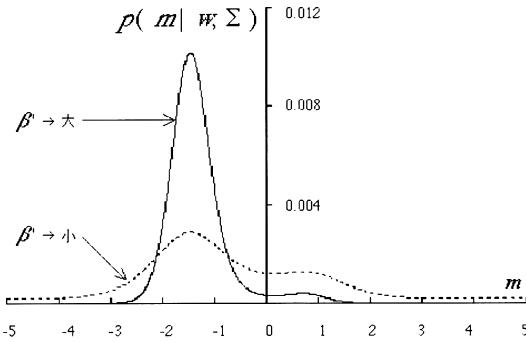


図 2 パラメータ m の条件付き確率密度関数
Fig. 2 The conditional probability density function of the parameter m .

$$p_{\beta'}(m|w, \Sigma) = \frac{p_{\beta'}(w, \phi)}{\int_m p_{\beta'}(w, \phi) dm} \quad (25)$$

は導出することができ、それは図 2 のようになる。

この結果から、シナプス結合荷重 $w = 1$ 、パラメータ $\Sigma = 0.2$ をもつ動径基底関数が与えられた条件のもとで累積 2 乗誤差関数 $E(w, \phi)$ を最小とするためには、パラメータ m を条件付き確率 $p_{\beta'}(m|w, \Sigma)$ を最大とする値に定めればよいことがわかる。また、もし同じ形質（パラメータ $w = 1, \Sigma = 0.2$ ）をもつ動径基底関数を一つ追加することができるなら、条件付き確率 $p_{\beta'}(m|w, \Sigma)$ を極大とするパラメータ m へ配置することが最も累積 2 乗誤差関数 $E(w, \phi)$ を小さくできることもわかる。

式 (25) を更に、シナプス結合荷重 w とパラメータ Σ で積分をとれば確率 $p_{\beta'}(m)$ が算出できる。そこで、教師信号を復元するために確率 $p_{\beta'}(m)$ を極大とするパラメータ m に動径基底関数を配置するような一撃アルゴリズムを考えることもできる。しかしながら、多次元の場合にはシナプス結合荷重 w とパラメータ Σ の積分が困難であることから、本研究では確率 $p_{\beta'}(m)$ を用いずに、条件付き確率を用いて逐次的にパラメータ m を求めていく方法を考える。

4. 動径基底関数の複製アルゴリズム

4.1 自由エネルギーの導出

一般に、式 (6) に従いパラメータ m_j^k を更新し続けると極小解にとらわれ、累積 2 乗誤差関数 $E(\mathbf{w}, \phi)$ の値を 0 にすることができないことがある。または、近似しようとしている非線形関数 $\eta(\mathbf{x})$ の複雑さに対し、

必要とされる動径基底関数が十分に存在していないときには、2 乗誤差関数 $E(\mathbf{w}, \phi)$ の値を 0 にすること 자체が不可能である。

ところで、確率的な要素や未知の教師信号などが存在しないものとするなら、すべての入力ベクトル \mathbf{x}_i ごとに動径基底関数を作成し、シナプス結合荷重が $w_i = \eta(\mathbf{x}_i)$ かつパラメータ $\Sigma_i \rightarrow 0$ であるときに、パラメータ m_i が \mathbf{x}_i となることで近似的に $E(\mathbf{w}, \phi) \approx 0$ とできる場合がある。ここで、0 は零行列を表す。もちろん、多くの問題ではすべての入力ベクトルについて動径基底関数を用意しなくても、このようなことが可能であるものと思われる。

そこで本研究では、累積 2 乗誤差関数 $E(\mathbf{w}, \phi)$ の値がある正数 $\epsilon > 0$ より大きな値に収束し、学習が収束したと判断されるときに、新たに必要な動径基底関数を追加する手法を提案する。ここで提案する手法では、前章で導出した確率密度関数を利用しているため、学習が収束した時点で得られている動径基底関数の一部の形質（シナプス結合荷重 w_j 、パラメータ Σ_j ）が新たに追加される動径基底関数がもつパラメータに引き継がれている。そのため、効率的に最も累積 2 乗誤差関数を小さくするパラメータ m に動径基底関数を追加していくことができる。なおかつ、最悪の場合にはすべての入力ベクトル \mathbf{x}_i をパラメータ m_i とする動径基底関数を作成することができる。そこで、この手法を CRBFN に組み入れたニューラルネットワークを複製・競合動径基底関数ネットワークと呼ぶこととする。

ところで、累積 2 乗誤差関数 $E(\mathbf{w}, \phi)$ の最小化は各入力ベクトル \mathbf{x}_i ごとに 2 乗誤差関数 $E(\mathbf{x}_i, \mathbf{w}, \phi)$ を最小化することに等価である。そこで、各入力ベクトル \mathbf{x}_i に依存した平均ベクトル $\mathbf{m}_{j[i]}$ を考える。そして、学習収束の時点で得られている第 j 番目の動径基底関数に着目すると、入力ベクトル \mathbf{x}_i の条件付き確率密度関数は

$$\begin{aligned} p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]}, \phi'_j, \phi''_j) \\ = Z_{\beta'}^{-1}(\mathbf{m}_j, \phi'_j, \phi''_j) \\ \times \exp\{-\beta'E(\mathbf{x}_i, \mathbf{m}_{j[i]}, \phi'_j, \phi''_j)\} \end{aligned} \quad (26)$$

と導出できる。ここで、パラメータ ϕ'_j は着目した第 j 番目の動径基底関数のシナプス結合荷重 w_j と共に分散行列 Σ_j の集合であり、パラメータ ϕ''_j は着目した第 j 番目の動径基底関数以外のシナプス結合荷重、共分散行列並びに平均ベクトルの集合である。以後は

記法の簡便のため、パラメータ ϕ'_j とパラメータ ϕ''_j は省略する。また、分配関数は

$$Z_{\beta'}(\mathbf{m}_j) = \sum_{i=1}^N \exp\{-\beta'E(\mathbf{x}_i, \mathbf{m}_{j[i]})\} \quad (27)$$

で定義される。

条件付き確率密度関数 $p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]})$ は、確率の正規化と 2 乗誤差関数 $E(\mathbf{x}_i, \mathbf{m}_{j[i]})$ の条件付き期待値

$$\langle E(\mathbf{m}_j) \rangle_{\beta'} = \sum_{i=1}^N p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]}) \times E(\mathbf{x}_i, \mathbf{m}_{j[i]}) \quad (28)$$

が一定となるという二つの制約のもとで、エントロピー

$$S_{\beta'}(\mathbf{m}_j) = -\frac{1}{\beta'} \sum_{i=1}^N p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]}) \times \log p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]}) \quad (29)$$

を最大にする確率密度関数として導出できる[10]。ここで、記号 $\langle \cdots \rangle_{\beta'}$ は $p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]})$ を掛けて \mathbf{x}_i に関する和をとる演算を表すものとする。このとき、自由エネルギーを

$$F_{\beta'}(\mathbf{m}_j) = -\frac{1}{\beta'} \log Z_{\beta'}(\mathbf{m}_j) \quad (30)$$

で定義すれば、

$$S_{\beta'}(\mathbf{m}_j) = -F_{\beta'}(\mathbf{m}_j) + \beta' \langle E(\mathbf{m}_j) \rangle_{\beta'} \quad (31)$$

と表すことができる。式(31)はエントロピー $S_{\beta'}(\mathbf{m}_j)$ を最大化する条件付き確率密度関数 $p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]})$ は、自由エネルギー $F_{\beta'}(\mathbf{m}_j)$ を最小化するものであることを示している。

このような自由エネルギーは、データのクラスタリングのための手法であるメルティング[11]においても同様に定義されている。メルティングとは、 $m_{j[i]} = \mathbf{x}_i$ ($i = 1, 2, \dots, N$)かつ β' が ∞ である初期状態から、徐々に β' を 0 へ近づけていきながら、パラメータ $m_{j[i]}$ を自由エネルギー $F_{\beta'}(\mathbf{m}_j)$ の最急降下方向に更新していくものである。その結果、パラメータ $m_{j[i]}$ は徐々に同じ値をとりはじめ、最終的に一つの値 $m_{j[i]} = m_j$ ($\forall i$) に収束する。

4.2 複製する位置の決定法

そこで、RC-RBFN ではパラメータ m_j^k の更新則

を式(6)の Δm_j^k の代わりに

$$\begin{aligned} \Delta_{\beta'} m_j^k &= -\epsilon \sum_{i=1}^N \frac{\partial F_{\beta'}(\mathbf{m}_j)}{\partial m_{j[i]}^k} \\ &= -\epsilon \sum_{i=1}^N \frac{\partial F_{\beta'}(\mathbf{m}_j)}{\partial Z_{\beta'}(\mathbf{m}_j)} \frac{\partial Z_{\beta'}(\mathbf{m}_j)}{\partial m_{j[i]}^k} \\ &= \sum_{i=1}^N p_{\beta'}(\mathbf{x}_i | \mathbf{m}_{j[i]}) \Delta m_{j[i]}^k \\ &= \langle \Delta m_j^k \rangle_{\beta'} \end{aligned} \quad (32)$$

で与えることとする。ここで、

$$\Delta m_{j[i]}^k = -\epsilon \frac{\partial E(\mathbf{x}_i, \mathbf{m}_{j[i]})}{\partial m_{j[i]}^k} \quad (33)$$

である。

特に $\beta' = 0$ であり、初期の状態が $m_{j[i]} = m_j$ ($\forall i$) である場合は

$$\Delta_0 m_j^k = \Delta m_j^k \quad (34)$$

であることが示される。この場合は、RC-RBFN のパラメータ m_j^k の更新則が従来の RBFN のパラメータ m_j^k の更新則そのものとなっていることがわかる。このとき、 $\beta' = 0$ で固定したままパラメータを $\Sigma_j \rightarrow 0$ にすると、 $\Delta_0 m_j^k = 0$ とするパラメータ $\mathbf{m}_{j[i]}$ は

$$\begin{aligned} &\sum_{i=1}^N \xi(\mathbf{x}_i, \mathbf{m}_{j[i]})(\mathbf{x}_i - \mathbf{m}_{j[i]}) \\ &\times \{\eta(\mathbf{x}_i) - s(\mathbf{x}_i, \mathbf{m}_{j[i]})\} = 0 \end{aligned} \quad (35)$$

を満たし、 $\mathbf{m}_{j[i]} = \mathbf{x}_i$ ($\forall i$) であることがわかる。つまり、教師入力信号がパラメータ \mathbf{m}_j の収束点として検出されることとなる。

得られた結果を確認するために、式(24)の教師信号から適当に 5 点選んだのが表 1 である。これを教師入力信号 x_i ($i = 1, 2, \dots, 5$) とする。式(32)において β' を 0、パラメータを Σ を徐々に小さくしたときに、 $\Delta_0 m_{j[i]} = 0$ となる点をプロットしたのが図 3 である。パラメータ $\Sigma \rightarrow 0$ において $m_{j[i]} = x_i$ ($i = 1, 2, \dots, 5$) へ収束していることがわかる。ただし、この

表 1 教師信号の値
Table 1 The values of teaching signals.

x_i	-4.00	-3.50	-1.00	0.50	3.00
$\eta(x_i)$	0.13	1.82	2.68	1.96	0.04

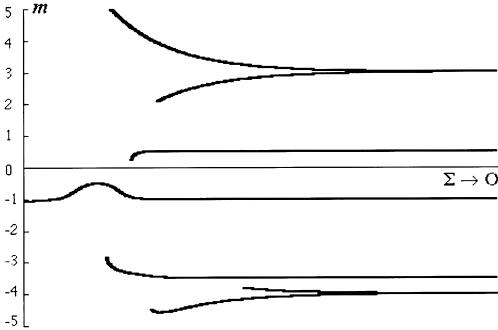


図 3 $\Delta_0 m = 0$ となるパラメータ m
Fig. 3 The parameter m satisfying $\Delta_0 m = 0$.

例では $j = k = 1$ であるため、添字の j と k は省略した。

あるいは、逆にパラメータ Σ_j を固定したまま $\beta' \rightarrow \infty$ にすると、 $\Delta_\infty m_j^k = 0$ とするパラメータ $\mathbf{m}_{j[i]}$ は

$$\sum_{i=1}^N \exp\{-\beta' E(\mathbf{x}_i, \mathbf{m}_{j[i]})\} \xi(\mathbf{x}_i, \mathbf{m}_{j[i]}) \\ \times (\mathbf{x}_i - \mathbf{m}_{j[i]}) \{\eta(\mathbf{x}_i) - s(\mathbf{x}_i, \mathbf{m}_{j[i]})\} = 0 \quad (36)$$

を満たし、 \mathbf{x}_i ($\forall i$) を含む任意の値となることがわかる。

これらの結果から、提案する RC-RBFN のパラメータ m_j^k の更新則 $\Delta_{\beta'} m_j^k$ では、 $\Delta_0 m_j^k$ で従来の RBFN のパラメータ m_j^k の更新則を実現し、更に、 $\Sigma_j \rightarrow 0$ とすればすべての入力ベクトル \mathbf{x}_i の第 k 要素 x_i^k ($\forall i$) をパラメータ m_j^k の安定な収束点として検出できることがわかる。あるいは、 $\Delta_\infty m_j^k$ とすることで、すべての入力ベクトル \mathbf{x}_i の第 k 要素 x_i^k ($\forall i$) を含む任意の値を安定な収束点とすることができます。

ここで、提案手法とゆう度解析との関係について示す。まず、次のような自由エネルギー

$$F_{\beta'} = -\frac{1}{\beta'} \log Z_{\beta'} \quad (37)$$

を考える。ただし、分配関数は

$$Z_{\beta'} = \sum_{i=1}^N \sum_{j=1}^M \exp\{-\beta' E(\mathbf{x}_i, \mathbf{m}_{j[i]})\} \quad (38)$$

で与えられる定数である。このとき、式 (30) は

$$F_{\beta'}(\mathbf{m}_j) - F_{\beta'} = -\frac{1}{\beta'} \log \frac{Z_{\beta'}(\mathbf{m}_j)}{Z_{\beta'}} \\ = -\frac{1}{\beta'} \log \sum_{i=1}^N p_{\beta'}(\mathbf{x}_i, \mathbf{m}_{j[i]}) \quad (39)$$

に変形することができる。ここで、確率密度関数 $p_{\beta'}(\mathbf{x}_i, \mathbf{m}_{j[i]})$ は

$$p_{\beta'}(\mathbf{x}_i, \mathbf{m}_{j[i]}) = Z_{\beta'}^{-1} \exp\{-\beta' E(\mathbf{x}_i, \mathbf{m}_{j[i]})\} \quad (40)$$

である。つまり、自由エネルギー $F_{\beta'}(\mathbf{m}_j)$ のパラメータ m_j^k に関する最小化は、対数ゆう度関数に関する最大化に等価であることを示すことができる。このような対数ゆう度関数と自由エネルギーとの関係は、EM アルゴリズムに関しては既に詳しい議論がされている [12]。

以上のことから、動径基底関数の複製を考慮した RC-RBFN の学習則を次のように提案する。

[RC-RBFN の学習則]

STEP 1. シナプス結合荷重 w_j を式 (8) のシナプス可塑性方程式により更新、パラメータ m_j^k を式 (32) の $\Delta_0 m_j^k$ により更新、パラメータ Σ_j は式 (7) により更新する。

STEP 2. 累積 2 乗誤差関数が $E(\mathbf{w}, \phi) \approx 0$ となつたら学習終了。ある正数 $\epsilon > 0$ より大きな値に収束したなら STEP 3. へいく。

STEP 3. 学習収束時に得られているすべての動径基底関数について、 β' を 0 から徐々に大きくしていくながら、式 (32) に従いパラメータ m_j^k を更新する。

STEP 4. 分岐により $\Delta_{\beta'} m_j^k = 0$ となる点が増えたとき、第 j 動径基底関数を第 p 動径基底関数として複製する。そのとき、シナプス結合荷重 w_p 、パラメータ Σ_p 並びにパラメータ m_p^n ($n \neq k$) は形質としてもとの第 j 動径基底関数のものを引き継ぎ、パラメータ m_p^k は新たに増えた点とする。STEP 1. へ戻る。

5. シミュレーション並びに考察

まず、本研究で新たに提案した動径基底関数の複製についてその有効性を確認する。そこで、教師信号を発生させる関数 $\eta(x)$ として

$$\eta(x) = \sum_{j=1}^N w_j \xi(x, \phi_j) \quad (41)$$

を考えた。ここでは、 $N = 3$ として $[w_1, w_2, w_3]^T = [2, 1, 1.5]^T$ 、 $[m_1, m_2, m_3]^T = [-4, -1, 4]^T$ 、 $[\Sigma_1, \Sigma_2, \Sigma_3]^T = [2^2, 0.5^2, 1^2]^T$ とした。初期に与えられる動径基底関数は一つであると仮定し、そのシナプス結合荷重を $w = 1.0$ 、パラメータを $m = 0$ 、 $\Sigma = 1.0$ とした。図 4(a) に教師信号と初期の動径基底関数を併

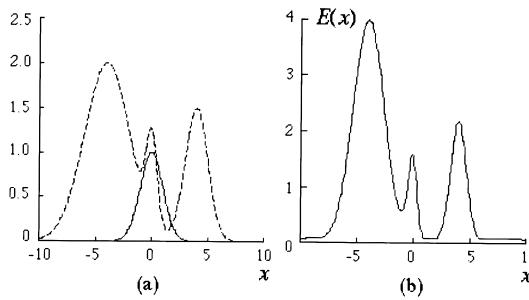


図 4 教師信号と初期の動径基底関数

Fig. 4 The distribution of the teaching signals and the initial radial basis function.

せて示す。図 4 (b) は初期の動径基底関数の出力を 0 と仮定したときの 2 乗誤差関数 $E(x)$ を示す。教師信号は $x \in [-10, 10]$ の間から等間隔に 201 個とり、 $(x_i, \eta(x_i)) i = 1, 2, \dots, 201$ で与えた。

このような教師信号と初期の動径基底関数が与えられたとき、先に提案した RC-RBFN に対する複製アルゴリズムを適用した結果を述べる。ただし、ここでは動径基底関数の追加による効果に着目するために、STEP 1.においてシナプス結合荷重 w とパラメータ Σ の更新は省略している。つまり、新たな安定な収束点が検出されると、そのパラメータ m に同じシナプス結合荷重 $w = 1$ とパラメータ $\Sigma = 1$ をもつ新たな動径基底関数を複製し、パラメータ m は更新されるがシナプス結合荷重 w とパラメータ Σ の更新をしない RC-RBFN に等価となっている。

学習に伴うパラメータ m の安定な収束点の変化を、横軸にパラメータ m をとり、縦軸に $\Delta F_{\beta'}(m) = m + \epsilon' \Delta_{\beta'} m$ をプロットしたものを図 5 に示す。ただし、定数 $\epsilon' = 100$ は図を見やすくするための係数である。安定な収束点は \bullet で示している。また、図 6 に学習による 2 乗誤差関数 $E(x)$ の変化を示す。

まず、図 4 (a) の初期の状態から $\beta' = 0$ として学習が始まると、初期の動径基底関数のパラメータ m は 0 から -3.99 へ移動する、同時にもう一つ安定な収束点が検出され $m = 4.03$ に新たな動径基底関数が複製される(図 5(a) 参照)。その結果、2 乗誤差関数は図 4 (b) から図 6 (a) のように変化する(以上 STEP 1. ~ STEP 4.)。更に、動径基底関数が二つとなった状態で学習を続けても、図 5 (b) のようになり学習が収束していることがわかる。そのときの、2 乗誤差関数は図 6 (b) であり、ほとんど図 6 (a) から変化してい

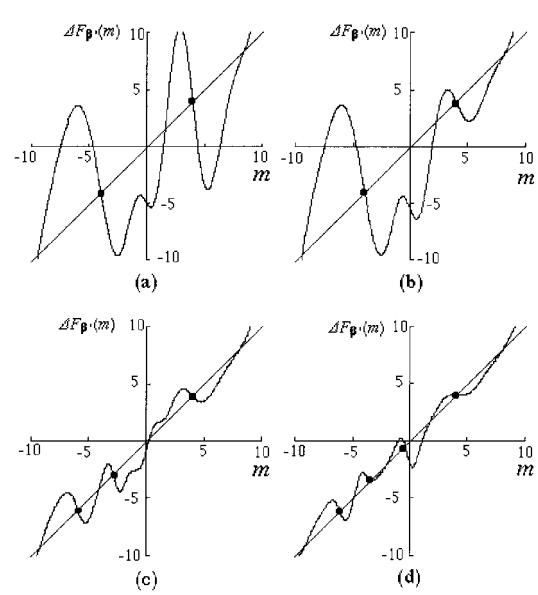
図 5 学習中のパラメータ m の安定な収束点

Fig. 5 The stable fixed point of parameter m during learning.

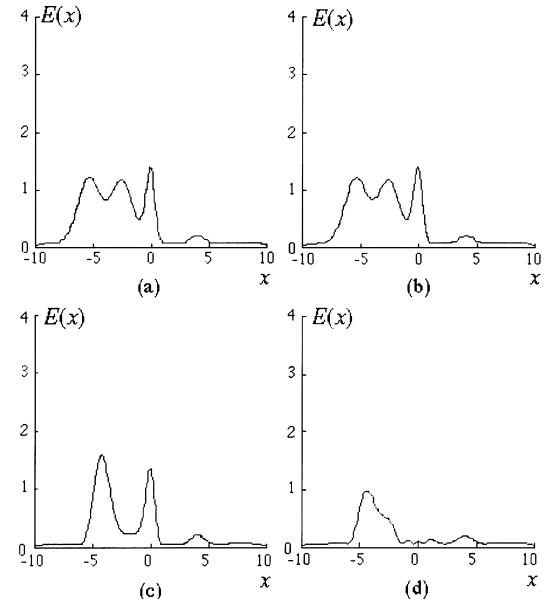


図 6 2 乗誤差関数の振舞い

Fig. 6 The behavior of square error function $E(x)$.

ない(以上 STEP 1. ~ STEP 2.).そこで、 β' を徐々に大きくしていながら二つのパラメータ m を更新しつつ、新たな安定な収束点が検出されると複製をす

る(図5(c)参照). パラメータ m の更新と動径基底関数の複製により, 図6(b)の $x = -5$ 附近にあった二つの極大値が効率良く減少していることが図6(c)からわかる(以上STEP 3.~STEP 4.). 更にこの学習則を繰り返すことにより, パラメータ m の更新並びに複製だけで2乗誤差関数が効率良く減少していることがわかる. この結果は, シナプス結合荷重を固定した状況で, 動径基底関数の複製アルゴリズムのみで得られたものである. 本論文で提案したアルゴリズムのSTEP 4においてシナプス結合荷重は固定されていることを考えると, 従来のRBFNにおいて動径基底関数を追加するアルゴリズムとして本手法を適用した場合でも, 同様の効果が期待できると考えられる.

次に, パラメータ m のみならず, シナプス結合荷重 w とパラメータ Σ の更新も行ったときに, 本研究で提案したRC-RBFNが従来の手法に比較して高速に学習ができる事を示す. そこで, 関数近似問題に対し学習時間やパラメータ数の点で優れているとされる最大値誤差選択法(Maximum Absolute Error selection method; 以下, MAE法)[13]との比較を行った. MAE法は累積2乗誤差関数の値がある正の値に収束したと判断されるとき, 2乗誤差関数 $E(x_i)$ が最大となる入力ベクトル x_i を新しい動径基底関数のパラメータ m として複製していく手法である. MAE法自体には動径基底関数を削除する手法が考慮されていないため, 初期の動径基底関数は一つで与えられることが多い.

そこで, 本研究でも初期に与える動径基底関数はRC-RBFNとMAE法とともに一つであるとして, 初期条件, 収束判定条件, 学習終了条件などの条件はすべて同じものを適用した. 計測したのは学習回数と学習時間である. また, より複雑な教師信号を発生させるために式(41)において, $N = 10$ として w_j, m_j, Σ_j ($j = 1, 2, \dots, 10$) は適当な乱数で決定した. 図7に作成された教師信号の様子を示す. 図8に学習回数に対する累積2乗誤差関数の値の変化を示す. シミュレーションに用いた機種はUltra SPARC 143 MHzであり, OSはSolaris2.5.1である. 学習時間はRC-RBFNで431秒, MAE法で1216秒であった. 教師信号を変えた実験では, 教師信号を発生させる関数を複雑にするほど, RC-RBFNはMAE法に比較して高速に学習が終了する傾向が見られた.

先にも述べたとおり, MAE法自体には動径基底関数を削除する手法が考慮されていない. そこで, 我々

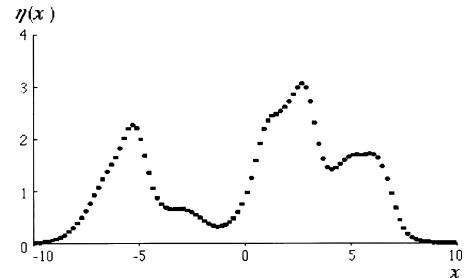


図7 教師信号の分布
Fig. 7 The distribution of the teaching signals.

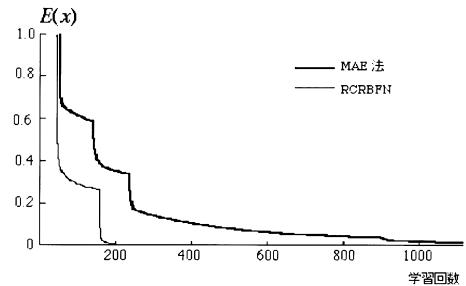


図8 累積2乗誤差関数の変化
Fig. 8 The change of the total square error function.

はCRBFNのシナプス結合荷重の更新則(式(8))をMAE法に組み込んだ実験も行った.しかし, この組合せでは動径基底関数の数が振動を起こすこともあり, 考察の対象とできなかったことを付記する.これは動径基底関数を2乗誤差関数の最大となる入力ベクトルに追加するアルゴリズム(MAE法)と, シナプス可塑性方程式により削除するアルゴリズム(適者生存型学習則)が全く無関係に行われていることに起因すると考えられる.これに比較してRC-RBFNが振動を抑えることができる原因是, 動径基底関数を追加するアルゴリズムが適者生存型学習則から得られる自由エネルギーを用いて導出されているため, 追加と削除が互いに相関し合うためであると思われる.しかし, これらのことと定式化し明確に記述することは困難であり, 今後に残された課題といえる.

以上の結果から, RC-RBFNは関数近似問題に適した従来のRBFNに, 効率良く動径基底関数を削除あるいは追加する能力を与えたニューラルネットワークであるといえる.

6. む す び

本論文では、シナプス可塑性方程式に関する考察から、必要な動径基底関数を効率的に追加する手法の提案をした。この手法では、学習が収束した時点で得られている動径基底関数の特性の一部を新たに追加される動径基底関数が引き継がれることとなる。そこで、本手法を競合動径基底関数ネットワークに組み合わせたニューラルネットワークを複製・競合動径基底関数ネットワークとして提案した。

シミュレーションでは、複製・競合動径基底関数ネットワークは複製により効率良く累積2乗誤差関数をより小さくでき、従来の動径基底関数ネットワークに比較して高速に学習できることが示された。

文 献

- [1] J. Park and I.W. Sandberg, "Universal approximation using radial basis function networks," *Neural Computation*, vol.3, pp.246–257, 1991.
- [2] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol.2, pp.148–153, 1989.
- [3] D.B. Fogel, "An information criterion for optimal neural network selection," *IEEE Trans. Neural Networks*, vol.2, pp.490–497, 1991.
- [4] 増井達也, 池谷浩彦, 藤井善行, "隠れユニットの分裂と合成による適切なサイズのニューラルネットワーク構造の獲得手法," *電学論C*, vol.113, no.10, pp.865–871, 1993.
- [5] Z. Wang, C.D. Massimo, M.T. Tham, and A.J. Morris, "A procedure for determining the topology for multilayer feedforward networks," *Neural Networks*, vol.7, pp.291–300, 1994.
- [6] 奥原浩之, 尾崎俊治, "適者生存型学習則を適用した競合動径基底関数ネットワーク," *信学論(D-II)*, vol.J80-D-II, no.12, pp.3191–3199, Dec. 1997.
- [7] B. Widrow and R. Winter, "Neural nets for adaptive filtering and adaptive pattern recognition," *IEEE Computer*, vol.21, pp.25–39, 1988.
- [8] 奥原浩之, 尾崎俊治, "Dale則を考慮したシナプス可塑性方程式の解析," *システム制御情報学会論文誌*, vol.8, no.12, pp.718–720, 1995.
- [9] 奥原浩之, 尾崎俊治, "パラメータ同定による単純系のポテンシャル形成," *信学技報*, NC97-15, June 1997.
- [10] H. Haken, "Synergetics - An Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology," Springer-Verlag, 1978.
- [11] Y. Wong, "Clustering data by melting," *Neural Computation*, vol.5, pp.89–104, 1993.
- [12] 上田修功, 中野良平, "確定的アニーリング EM アルゴリズムとその確率ニューラルネットワークへの応用," *信学技報*, NC95-10, May 1995.
- [13] R. Katayama, M. Watanabe, K. Kuwata, Y. Kajitani, and Y. Nishida, "Performance evaluation of self generating radial basis function for function approximation," *Proc. of IJCNN '93 Congress Nagoya*, vol.2, pp.471–474, 1993.

付 錄

ここで、本論文で展開した議論を従来の RBFN に対して適用すれば、動径基底関数を効率的に追加することが可能な学習則が導けること端的に述べる。式(5)で与えられる Delta ルールに基づいて導出された RBFN のシナプス結合荷重の更新則を内的自然増加率 $\alpha_j(\phi)$ 、並びに競争係数 $\gamma_{jk}(\phi)$ を用いて記述すると

$$\begin{aligned}\Delta w_j &= -\epsilon \frac{\partial E(\mathbf{w}, \phi)}{\partial w_j} \\ &= \epsilon \left\{ \sum_{i=1}^N \eta(\mathbf{x}_i) \xi(\mathbf{x}_i, \phi_j) \right. \\ &\quad \left. - \sum_{h=1}^M w_h \sum_{i=1}^N \xi(\mathbf{x}_i, \phi_j) \xi(\mathbf{x}_i, \phi_h) \right\} \\ &= \epsilon \left(\alpha_j(\phi) - \sum_{k=1}^M \gamma_{jk}(\phi) w_k \right) \quad (\text{A.1})\end{aligned}$$

となることがわかる。これは積分条件

$$\frac{\partial}{\partial w_k} \frac{dw_j}{dt} = \frac{\partial}{\partial w_j} \frac{dw_k}{dt} \quad (\text{A.2})$$

を満たすことから、ポテンシャル

$$\begin{aligned}V(\mathbf{w}, \phi) &= - \sum_{j=1}^M \left\{ \alpha_j(\phi) w_j \right. \\ &\quad \left. - \sum_{k \neq j}^M \gamma_{jk}(\phi) w_j w_k - \frac{\gamma_{jj}(\phi)}{2} w_j^2 \right\} \quad (\text{A.3})\end{aligned}$$

を考えることができる。ただし、 ϵ は省略した。この結果、

$$E(\mathbf{w}, \phi) = V(\mathbf{w}, \phi) + \frac{1}{2} \sum_{i=1}^N \eta^2(\mathbf{x}_i) \quad (\text{A.4})$$

であることが示されるので、やはり累積2乗誤差関数 $E(\mathbf{w}, \phi)$ の最小化はポテンシャル $V(\mathbf{w}, \phi)$ の最小化と等価であることがわかる。その結果、学習終了時に RBFN の平均ベクトル、共分散行列とシナプス結合荷

重が満たす同時確率密度 $p_\beta(\mathbf{w}, \phi)$ を考えることができ、自由エネルギーを定義することで式(32)に相当する平均ベクトルの更新則が導出できる。詳しい展開は冗長なため省略する。

もちろん、式(5)で与えられる RBFN のシナプス結合荷重の更新則がそのまま積分条件を満たすことから、ポテンシャル $V(\mathbf{w}, \phi)$ を導出せずに、累積 2 乗誤差関数 $E(\mathbf{w}, \phi)$ をポテンシャルとみなしても同様の議論が可能である。しかしながら、式(5)とそれから得られる追加アルゴリズムを利用した学習則では動径基底関数を削除する効果は望めない。そこで、本論文では適者生存型学習則に対する動径基底関数の複製について議論を展開した。

(平成 10 年 4 月 24 日受付、10 月 30 日再受付)

奥原 浩之（正員）

平 8 広島大大学院工学研究科システム工学専攻博士課程後期了。同年、広島大学工学部助手。平 10 広島県立大学経営学部講師。博士（工学）。協同現象の数理に興味をもち、自律分散システムの研究に従事。システム制御情報学会、日本 OR 学会各会員。

佐々木浩二

平 8 広島大・工・第二類経営卒。現在、広島大大学院工学研究科システム工学専攻博士課程前期在学中。ニューラルネットワークの応用に関する研究に従事。

尾崎 俊治（正員）

昭 45 京都大大学院博士課程了。同年広島大学工学部講師、同年助教授、昭 61 教授。工博。信頼性工学及びシステム工学の研究に従事。昭 45~52 マンチェスター大学研究員。昭 45~47 南カルフォルニア大学研究員。日本 OR 学会第 4 回大西文献賞授賞。著書「応用確率論」、「Stochastic System Reliability Modeling」など。情報処理学会、日本 OR 学会各会員。