

論文

競合連想ネットの漸近最適性と非線形関数の逐次学習への応用

黒木 秀一[†]

Asymptotic Optimality of Competitive Associative Nets and Its Application to Incremental Learning of Nonlinear Functions

Shuichi KUROGI[†]

あらまし 競合連想ネット CAN2 は競合ネットと連想ネットの機能を用いて非線形関数を逐次的に学習し区分的に線形近似するニューラルネットである。このネットの有効性は関数近似、制御、降水量推定などへの応用において示されているが、その学習法は基本的にこう配法に基づく競合学習を含んでおり、局所解問題をもつ。この問題を回避するために、本論文ではまずネットを構成するユニット数が非常に多い場合を考え、近似すべき関数の平均 2 乗誤差を最小化するための最適性の条件、すなわち漸近最適条件を導く。この条件は各時刻の荷重配置が最適解に近いかどうかを判断するために用いることができると同時に、最適解により近い荷重配置を示唆することができる。そこでこの条件をこう配法による学習に組み入れる。すなわちこう配法により得られた荷重配置が最適解に近くないと判断されるときはあるユニットの荷重を最適解に近い荷重配置になるように再初期化する学習法を構成する。最後にこの学習法をいくつかのベンチマーク関数に適用する数値実験を通して本手法の有効性を確認するとともに、BPN (Back-Propagation Net), RBFN (Radial Basis Function Net) 及び SVR (Support Vector Regression) の実験結果と比較し、本手法を用いる CAN2 の関数近似性能の高さを示す。

キーワード 競合連想ネット, 漸近最適性, 競合学習, 局所解問題, 関数近似

1. まえがき

競合連想ネット CAN2 (Competitive Associative Net 2) は従来の競合ネットと連想ネットの機能を結合して非線形関数を逐次的に学習し区分的に線形近似するために導入したニューラルネットである [1]。ここで関数を区分的に線形近似 (あるいは内挿) する手法としては Dyn らの手法 [2] や Near-Neighbor 回帰法 [3] などがあるが、逐次学習を行う区分的線形近似手法は競合連想ネット以外は見当たらない。一方、関数近似にしばしば応用されるニューラルネットの代表的な手法である BPN (Back-Propagation Net), RBFN (Radial Basis Function Net) [4], SVR (Support Vector Regression) [5] などは通常、非線形連続な入出力関数を実現するネットワーク構造をもち逐次的な非線形最適化手法であるこう配法や 2 次元最適化法などを使用するが、CAN2 はそれらとは異なるネットワーク構造

と逐次最適化法すなわち区分的線形関数を実現する競合連想ネット構造と競合学習と連想記憶 (線形最小 2 乗法) を組み合わせた学習法を採用している。CAN2 の定量的解析 [6] においては CAN2 が通常の BPN よりも優れた学習性能や関数近似能力をもつことが示され、更に非線形あるいは時変のプラントのモデル切替制御 [7]~[9]、非線形性が強く因果関係が不明確な降水量推定 [10] などへの応用においてはその有用性が確かめられている。CAN2 の学習能力と区分的線形近似能力は、制御への応用においては非線形系や時変系の複数の動作点における複数の線形モデル (線形の入出力関係) を一つのニューラルネットで学習可能とするとともに、各時刻で一つの線形モデルを選択して関数近似するので従来の効率の良い線形制御則 (例えば一般化予測制御則) の使用を可能にしている。更に CAN2 の関数近似能力の高さは、2000 年度電子情報通信学会総合大会シンポジウム・降水量推定コンテストにおいて提出された種々の手法の中で CAN2 を用いた降水量推定 [10] が第 2 位の成績を得たことからわかる。

一方、従来より競合ネットはパターン認識やベクト

[†]九州工業大学工学部, 北九州市
Faculty of Engineering, Kyushu Institute of Technology,
Kitakyushu-shi, 804-8550 Japan

ル量子化 (VQ; Vector Quantization) の分野などに
 応用されているが, その学習すなわち競合学習は一般
 にはこう配法に基づいて局所解を求めるため, 学習結
 果が荷重ベクトルの初期値に依存する, あまり利用さ
 れない荷重ベクトルが生じる, などのいわゆる局所解
 問題を本質的にもっている. しかし最近, VQ の研究
 においては荷重ベクトル数が十分大きいときの最適性
 の条件 (漸近最適条件という) が導出され [11], その
 条件を用いた学習アルゴリズム [12], [13] は荷重ベクト
 ル数が比較的少数の場合にも局所解を回避するのに有
 効であることが確かめられている.

本論文は VQ の研究を参考に CAN2 の漸近最適条
 件を導出するとともに, より良い性能をもつ学習アル
 ゴリズムを構築することを目的とする. 以下, 2. で
 CAN2 とその一般的な学習手順を示し, 3. で漸近最
 適条件を導く. 4. では漸近最適条件を逐次的に計算
 する方法を示すとともに学習手順に組み込み, 5. で
 は数値実験結果を通して本手法の有効性を確認すると
 ともに他手法との比較を行い本手法の関数近似性能の
 高さを示す.

2. CAN2 と学習手順

離散時間 $t = 1, 2, \dots$ において k 次元ベクトル
 $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_k(t))^T \in \mathbb{R}^{k \times 1}$ とスカラー値
 $y(t) \in \mathbb{R}$ を入出力とする関数

$$y(t) \triangleq f(\mathbf{x}(t)) + d(t) \quad (1)$$

を考える. ここで $d(t) \in \mathbb{R}$ は平均 $\mu_d = 0$, 分散 σ_d^2
 の観測雑音とする. 以下では誤解が生じないときは簡
 単のため $y = f(\mathbf{x}) + d$ のように時刻 t を省いて表す.

CAN2 は荷重ベクトル $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{ik})^T \in \mathbb{R}^{k \times 1}$
 と連想行列 $\mathbf{M}_i = (M_{i0}, M_{i1}, \dots, M_{ik}) \in \mathbb{R}^{1 \times (k+1)}$ をもつユニットを N 個 ($i = 1, 2, \dots, N$)
 用いて (図 1 参照),

$$\hat{y} = \hat{y}_c \triangleq \mathbf{M}_c \tilde{\mathbf{x}} \quad (2)$$

により関数 $y = f(\mathbf{x})$ の近似を行う. ここで本論文で
 はベクトル $\tilde{\boldsymbol{\xi}} = (\xi_0, \xi_1, \dots, \xi_k)^T$ は $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_k)^T$
 の第 1 要素の前にスカラー ξ_0 を付加したベクトル
 を表し, 上式の $\tilde{\mathbf{x}}$ では $x_0 = 1$ とした. また,

$$c \triangleq \underset{j=1,2,\dots,N}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{w}_j\| \quad (3)$$

は競合により選択されるユニットの番号を表し, あ

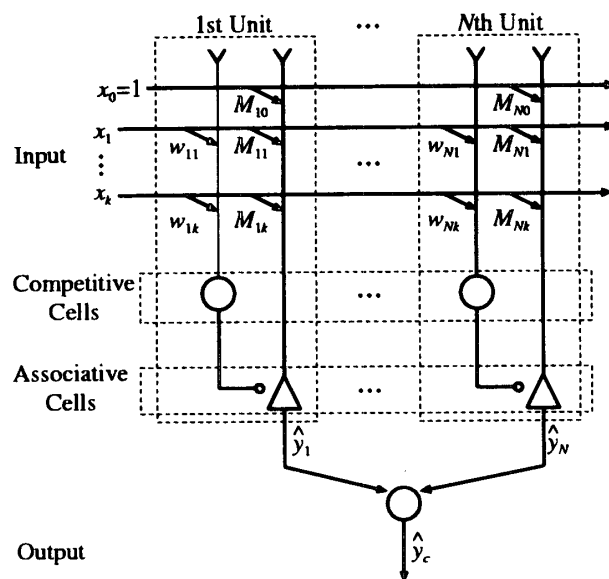


図 1 CAN2 の構造
 Fig. 1 Schematic diagram of the CAN2.

るユニットがこの競合に勝って出力することをそのユ
 ニットが発火するともいう. 更にこの関数近似は入力
 空間 $V = \mathbb{R}^k$ を $i = 1, 2, \dots, N$ のボロノイ領域 (ま
 たはディリクレ領域ともいう)

$$V_i \triangleq \{\mathbf{x} \mid i = \underset{j=1,2,\dots,N}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{w}_j\|\} \quad (4)$$

に分割して, 関数 $y = f(\mathbf{x})$ の区分的線形近似を行う
 ことを意味する. なおボロノイ領域に分割することを
 ディリクレ分割ともいい, 分割されたボロノイ領域は
 凸多面体になり, 更に有界な凸多面体はポリトープと
 呼ばれる.

CAN2 の学習は近似誤差 $e(\mathbf{x}) \triangleq y - \hat{y}$ の 2 乗の期
 待値を表すネットのエネルギー

$$E \triangleq \int_V \|e(\mathbf{x})\|^2 p(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^N \int_{V_i} \|e_i(\mathbf{x})\|^2 p(\mathbf{x}) d\mathbf{x} = \sum_{i=1}^N E_i \quad (5)$$

を逐次的に最小化することを目的とする. ここで $p(\mathbf{x})$
 は入力 \mathbf{x} の確率密度, $e_i(\mathbf{x}) \triangleq y - \hat{y}_i$ は第 i ユニ
 ヲットの近似誤差, $E_i \triangleq \int_{V_i} \|e_i(\mathbf{x})\|^2 p(\mathbf{x}) d\mathbf{x}$ は第 i ユニ
 ヲットのエネルギーを表す. なお第 i ユニットの 2 乗誤差
 の期待値は E_i ではなく $E_i / \int_{V_i} p(\mathbf{x}) d\mathbf{x}$ であること
 を注意しておく.

本論文で考える CAN2 の学習は表 1 の (i)~(iii) の
 処理からなる. 従来の CAN2 の学習も, 応用に応じ

表 1 CAN2 の学習手順
Table 1 Learning procedure for the CAN2.

| |
|--|
| (i) 各時刻において式 (3) の競合に勝つ第 c ユニットの求め、その競合ベクトル w_c を用いて再初期化条件を調べる。もし条件が満たされなければ (ii) に進む。そうでなければ最も非効率な第 l ユニットの探して、その M_l と w_l に初期値を代入した後、このユニットを第 c ユニットとし (これを再初期化という)、(ii) に進む。 |
| (ii) 第 c ユニットのエネルギー E_c を最小にするため、逐次的線形最小 2 乗法を用いて M_c を更新する。 |
| (iii) エネルギー E のこう配 $\partial E/\partial w_c$ の反対方向に w_c を微小移動させる。 |

て細かな違いはあるが、これとほぼ同様である。(ii) と (iii) の詳細は付録 1. に示すが、(iii) は一般にはディリクレ分割の局所最適解を求めるためのこう配法である。そこで (i) では w_c を用いて再初期化条件を調べるが、この条件はその時刻のディリクレ分割が最適解の近傍にないことを判断するためのものである。そこでこの条件が満たされれば最適解に近づけるための再初期化処理を行う。以下ではこの再初期化条件をユニット数 N が十分大きい場合の最適解が満たす条件 (漸近最適条件) に基づいて検討する。

3. CAN2 の漸近最適条件

確率密度関数 $p(\mathbf{x})$ は十分滑らかであり; $p(\mathbf{x}) \neq 0$ となる領域に十分多くの荷重ベクトル w_i ($i = 1, 2, \dots, N$) が配置されて各 V_i は十分小さくなり、 $p(\mathbf{x})$ は各 V_i 内で一定値 p_i で近似できるとする。更に学習すべき関数 $f(\mathbf{x})$ の 2 次導関数 $\mathbf{A} \triangleq [\partial^2 f/(\partial x_m \partial x_n)] \in \mathbb{R}^{k \times k}$ は連続であり、出力 $y = f(\mathbf{x}) + d$ は各 V_i 内で次式の 2 次関数により近似でき、

$$y \approx \tilde{\mathbf{a}}_i^T \tilde{\mathbf{x}} + \mathbf{x}^T \mathbf{A}_i \mathbf{x} + d, \quad (6)$$

更に隣接する \mathbf{A}_i は緩やかに変化するとする。ここで $\tilde{\mathbf{a}}_i = (a_{i0}, a_{i1}, \dots, a_{ik})^T \in \mathbb{R}^{(k+1) \times 1}$ は定数ベクトル、 $\mathbf{A}_i \in \mathbb{R}^{k \times k}$ は一般性を失うことなく実対称定数行列とする。

3.1 M_i の最適化と量子化係数

式 (5) と式 (6) よりポロノイ領域 V_i のエネルギーは

$$E_i = p_i \int_{V_i} \left((\tilde{\mathbf{a}}_i^T - \mathbf{M}_i) \tilde{\mathbf{x}} + \mathbf{x}^T \mathbf{A}_i \mathbf{x} + d \right)^2 d\mathbf{x} \quad (7)$$

となり、最適な $\mathbf{M}_i = \mathbf{M}_i^*$ はこの E_i を最小化することにより求まるが、この解は V_i の形状に依存する。

したがって全エネルギー E を最小化する \mathbf{M}_i^* を求めるには更に V_i の最適形状も求める必要があるが、この求解はかなり困難であり、既に多くの性質が解明されている VQ の分野においても一般には求まっていない (入力空間が 1 次元と 2 次元の場合はそれぞれ線分と正六角形になることは解明されている)。そこで VQ の分野ではこの問題を直接解くことは避け、ある性質をもつ最適な V_i の存在を仮定して漸近最適条件を導いている。そこで本論文でも同様の検討を以下のように行う。

まず非有界なポロノイ領域 V_i は $p(\mathbf{x}) \neq 0$ の領域 $V_{p(\mathbf{x}) \neq 0}$ と $p(\mathbf{x}) = 0$ の領域 $V_{p(\mathbf{x}) = 0}$ の境界に現れるが、ユニット数 N が十分大きいときはそれらの領域のエネルギー E_i の全体のエネルギー E に対する割合は非常に小さくできるので以下の議論では無視するとともに、有界なポロノイ領域 V_i をポリトープ (有界凸多面体) H_i で表記する。さて E の最小化を行いながら N を十分大きくしていくと、領域 $V_{p(\mathbf{x}) \neq 0}$ のある点 $\mathbf{x} = w_i$ における H_i の最適形状は N の増加にほとんど影響されなくなると考えられる。これは関数 $y = f(\mathbf{x})$ は緩やかに変化すると仮定しているので、 $V_{p(\mathbf{x}) \neq 0}$ に荷重ベクトル w_i ($i = 1, 2, \dots, N$) を十分多く配置すると、ある点 w_i と近隣の点 $w_j = w_i + \Delta w_{ij}$ における式 (6) の関数のパラメータ値はほとんど等しくなると同時に、 w_i から十分遠くにある点の関数のパラメータ値は H_i の最適形状にほとんど影響を与えないからである。そこで H_i の形状を保ったまま N を大きくして H_i の体積のみを小さくしていく状況を考える。このときこの点 w_i における正規化 2 乗誤差を次式で定義する。

$$I(H_i) \triangleq \frac{\int_{H_i} \|e_i^*(\mathbf{x})\|^2 d\mathbf{x}}{v_i^{1+r/k}}, \quad (8)$$

ここで $v_i = \int_{H_i} d\mathbf{x}$ は H_i の体積、 $e_i^*(\mathbf{x})$ は雑音 d を 0 とし \mathbf{M}_i をその最適値 \mathbf{M}_i^* としたときの近似誤差

$$e_i^*(\mathbf{x}) \triangleq (\tilde{\mathbf{a}}_i^T - \mathbf{M}_i^*) \tilde{\mathbf{x}} + \mathbf{x}^T \mathbf{A}_i \mathbf{x} \quad (9)$$

である。また r は x_j ($j = 1, 2, \dots, k$) の多項式 $\|e_i^*(\mathbf{x})\|^2$ の最高次数すなわち $r = 4$ とする。ここで式 (9) の代わりに $e_i^*(\mathbf{x}) = \mathbf{x} - w_i$ を式 (8) に用いると、その $I(H_i)$ は VQ における正規化慣性の定義になり、任意の $\xi > 0$ と任意のポリトープ H_i に対し、

$$I(\xi H_i) = I(H_i) \quad (10)$$

が成立することを示せるが、式 (9) を用いる場合は任意の H_i について式 (10) は成立しない。検討の結果、本研究では式 (6) の \mathbf{A}_i の固有ベクトルに依存して定義される k 次元直方体 R_i 及び k 次元楕円体 B_i に対してこの性質を得ることができた (付録 2. 及び 3. 参照)。そこでエネルギー E を最小化する最適ポリトープ H_i^* においてもこの性質が成立すると仮定し、そのときの正規化 2 乗誤差によりこの点 \mathbf{w}_i の量子化係数

$$C_i \triangleq C(\mathbf{w}_i) \triangleq I(H_i^*) \quad (11)$$

を定義する。なお $I(R_i)$ と $I(B_i)$ は式 (A.10) で定義される非線形性の強さを表す γ_i (\mathbf{A}_i の固有値の絶対値の積) の $2/k$ 乗に比例する。したがって量子化係数 C_i も同様に非線形性の強さ $\gamma_i^{2/k}$ に比例するのではないかと考えられる。

3.2 \mathbf{w}_i の最適化と漸近最適条件

量子化係数 C_i を用いると $E_i = C_i p_i v_i^{1+4/k} + \sigma_d^2 p_i v_i$ となり、 $E = \sum_{i=1}^N E_i$ は

$$\begin{aligned} E &= \sum_{i=1}^N C_i p_i v_i^{1+4/k} + \sigma_d^2 \dots \dots \dots \\ &\geq N^{-4/k} \left(\sum_{i=1}^N \left(C_i p_i \right)^{\frac{1}{1+4/k}} v_i \right)^{1+4/k} + \sigma_d^2 \\ &= N^{-4/k} \|C(\mathbf{x})p(\mathbf{x})\|_{\frac{1}{1+4/k}} + \sigma_d^2 \quad (12) \end{aligned}$$

となる。ここで $\|g(\mathbf{x})\|_\alpha = (\int_V |g(\mathbf{x})|^\alpha d\mathbf{x})^{1/\alpha}$ を表し、 $\|C(\mathbf{x})p(\mathbf{x})\|_{\frac{1}{1+4/k}}$ は与えられた $f(\mathbf{x})$ と $p(\mathbf{x})$ に対し定数になる。式 (12) の最下段の等号は N が十分大きいとき成立し、第 2 段の不等式は Hölder の不等式であり、等号が成立するのは

$$\alpha_i \triangleq C_i p_i v_i^{1+4/k} = \text{一定} \quad (i = 1, 2, \dots, N) \quad (13)$$

となる時のみであり、この式が漸近最適条件を表す。換言すると、すべての i に対して v_i が $(C_i p_i)^{1/(1+4/k)}$ に反比例するときのみ E を最小化する \mathbf{w}_i の最適配置が得られることを意味する。また式 (12) の最下段の式は E の最小値であり、その第 1 項は N の増加に対して $N^{-4/k}$ に比例しながら減少することを示す。

4. 漸近最適条件の逐次計算

漸近最適条件を逐次的に計算するために、まず忘却

付き発火回数 \hat{n}_i と忘却付き 2 乗誤差和 \hat{S}_i を各時刻において

$$\hat{n}_i := \begin{cases} \eta \hat{n}_i + 1 & (i = c) \\ \eta \hat{n}_i & (i \neq c) \end{cases} \quad (14)$$

$$\hat{S}_i := \begin{cases} \eta \hat{S}_i + \|e_i(\mathbf{x})\|^2 & (i = c) \\ \eta \hat{S}_i & (i \neq c) \end{cases} \quad (15)$$

により計算する。ここで $\eta = \exp(-1/\tau_\eta)$ は忘却係数、 τ_η は忘却の時定数、及び $:=$ は代入を表す。すべての $i = 1, 2, \dots, N$ について \mathbf{w}_i と \mathbf{M}_i が一定のまま時刻 t が十分長い時間区間 $[t_0, t_0 + n - 1]$ を経過し、かつその区間で第 i ユニットがそれぞれ平均時間間隔 n/n_i ($\ll \tau_\eta$) で発火するとすると、 \hat{n}_i の収束値は

$$\hat{n}_i \approx \frac{1}{1-\eta} \approx \frac{\tau_\eta n_i}{n} \approx \frac{\hat{n} n_i}{n} \quad (16)$$

と近似できる。ここでこの時間区間における第 i ユニットの発火回数 n_i は $n = \sum_{i=1}^N n_i$ を満たすので

$$\hat{n} \triangleq \sum_{i=1}^N \hat{n}_i \approx \sum_{i=1}^N \frac{\tau_\eta n_i}{n} = \tau_\eta \quad (17)$$

を式 (16) 中の τ_η と置き換えた。これにより忘却係数 η が 1 のときにも \hat{n}_i を見積もることができることになる。一方、式 (15) で $\|e_i(\mathbf{x})\|^2$ をその期待値 $E_i / \int_V p(\mathbf{x}) d\mathbf{x} \approx \alpha_i / (p_i v_i) + \sigma_d^2$ に置き換えた量が時間間隔 n/n_i ごとに加算されていくとすると、 \hat{S}_i の収束値は

$$\hat{S}_i \approx \frac{\alpha_i / (p_i v_i) + \sigma_d^2}{1-\eta} \approx \hat{n} \alpha_i + \sigma_d^2 \hat{n}_i \quad (18)$$

となる。ここで $n_i \approx n p_i v_i$ を用いた。すると α_i が一定値 α になるという漸近最適条件が成立するとして、上の式を $i = 1, 2, \dots, N$ について連立させると α と σ_d^2 の最小 2 乗解 $\hat{\alpha}$ と $\hat{\sigma}_d^2$ は次式で与えられる。

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\sigma}_d^2 \end{bmatrix} = \begin{bmatrix} \hat{n} & \hat{n}_1 \\ \dots & \dots \\ \hat{n} & \hat{n}_N \end{bmatrix}^+ \begin{bmatrix} \hat{S}_1 \\ \dots \\ \hat{S}_N \end{bmatrix}, \quad (19)$$

ここで行列 X の一般化逆行列を X^+ で表す。もし \mathbf{w}_i の最適配置が達成されていないと

$$\hat{\alpha}_i \triangleq (\hat{S}_i - \hat{\sigma}_d^2 \hat{n}_i) / \hat{n} \quad (20)$$

は一定ではなくなる。特に第 c ユニットの $\hat{\alpha}_c$ が $\hat{\alpha}_i$ ($i = 1, 2, \dots, N$) の平均 $(\hat{\alpha}_i) \triangleq \sum_{i=1}^N \alpha_i / N$ よりも

かなり大きいとき、すなわち 1 より十分大きな定数 θ_α に対して

$$\hat{\alpha}_c > \theta_\alpha \langle \hat{\alpha}_i \rangle \quad (21)$$

となるときは、 \mathbf{w}_i ($i = 1, 2, \dots, N$) の配置が最適解の近傍にないと判断できる。更にこの条件は $\hat{\alpha}_c \approx C_c p_c v_c^{1+4/k}$ が大きすぎることを意味するので、この領域 V_c の中に他のユニットを補充すれば v_c (したがって $\hat{\alpha}_c$) が小さくなり、最適解に近づくと考えられる。また $\hat{\alpha}_i$ が最小のユニットのデータを初期化してこの領域に補充すれば α_i の均一化と整合する。そこで表 1 (i) の再初期化条件として式 (21) を用い、非効率な第 l ユニットとして $\hat{\alpha}_i$ が最小のユニットを用いる。なお学習したデータ数が少ないときや \mathbf{M}_i や \mathbf{w}_i が変化した場合には上で仮定した統計的条件が満足されなくなり $\hat{\alpha}$ や $\hat{\sigma}_d^2$ の信頼性がなくなる。そこで初期時刻または最後に再初期化が行われた時刻 t_{RI} から十分長い時間 T_{RI} 経過した後、すなわち

$$t \geq t_{RI} + T_{RI} \quad (22)$$

に上記の再初期化条件の計算をするようにした。

5. 数値実験

5.1 準備

提案する学習アルゴリズムの学習性能を次の関数について検証した (図 2 及び [1], [2] 参照)。

$$f_1(\mathbf{x}) = \frac{\tanh(9x_2 - 9x_1) + 1}{9},$$

$$f_2(\mathbf{x}) = \frac{3}{4} \exp\left(-\frac{(9x_1 - 2)^2 + (9x_2 - 2)^2}{4}\right) + \frac{3}{4} \exp\left(-\frac{(9x_1 + 1)^2}{49} - \frac{(9x_2 + 1)^2}{10}\right) + \frac{1}{2} \exp\left(-\frac{(9x_1 - 7)^2 + (9x_2 - 3)^2}{4}\right) - \frac{1}{5} \exp(-(9x_1 - 4)^2 - (9x_2 - 7)^2),$$

$$f_3(\mathbf{x}) = \begin{cases} 1 & (x_2 - \xi > 1/2) \\ 2(x_2 - \xi) & (0 \leq x_2 - \xi \leq 1/2) \\ (\cos(4\pi r) + 1)/2 & (r \leq 1/4) \\ 0 & (\text{その他}) \end{cases}$$

$$\xi = 2.1x_1 - 0.1; r = \sqrt{(\xi - 3/2)^2 + (x_2 - 1/2)^2},$$

$$f_4(\mathbf{x}) = \begin{cases} 0 & ((x_1 - 0.5)(x_2 - 0.5) \geq 0) \\ 1 & (\text{その他}) \end{cases}$$

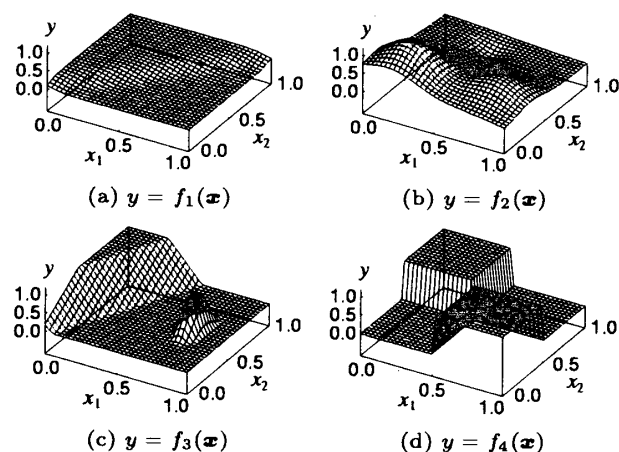


図 2 ベンチマーク関数
Fig. 2 Benchmark functions.

ここで関数 f_4 は $x_1 = 0.5$ と $x_2 = 0.5$ において不連続なのでその 2 次導関数も不連続であり漸近最適性の成立条件を局所的に満たさないことが容易にわかるが、関数 f_3 もその小山と平地の部分 (図 2 (c) 参照) がそれぞれ $f_3(\mathbf{x}) = (\cos(4\pi r) + 1)/2$ と $f_3(\mathbf{x}) = 0$ で定義されており、それらの境界では 2 次導関数が不連続であることを注意しておく。

各関数の訓練データ $y = f_i(\mathbf{x}) + d$ はベクトル \mathbf{x} を単位正方形 $S = \{\mathbf{x} = (x_1, x_2) \mid x_1, x_2 \in [0, 1]\}$ から各時刻 $t = 1, 2, \dots, 5 \times 10^7$ においてランダムに選んで $f_i(\mathbf{x})$ を計算したあと ± 0.01 の範囲の一樣乱数を加えて生成した。この訓練データの学習後、 S 上の 33×33 の格子点上の関数値 $y = f_i(\mathbf{x})$ (図 2 参照) とネットによる推定値 \hat{y} との平均 2 乗誤差 (MSE) を用いてネットを評価した。

比較のため、再初期化を行わない CAN2 (以下これを CAN2-OR (CAN2 withOut Reinitialization) と表す)、BPN、RBFN 及び SVR についても同様の実験を行った。BPN の学習係数は 0.07、慣性モーメントは 0.5 とした (学習手順を工夫すると以下の結果は少し改善されるがその詳細は [6] を参照のこと)。RBFN の出力 $g(\mathbf{x}) = \sum_{i=1}^N c_i \exp(-\|\mathbf{x} - \mathbf{w}_i\|^2/r_i^2)$ を決定するパラメータのうち基底関数の幅 r_i は、文献 [4], [14] 等で報告されているように、こう配法で r_i を探索するよりも単一の $r = r_i$ を使う方が良い結果を与えたので、 $r = 0.04, 0.06, \dots, 0.20$ 及び $N = 5, 50, 100, 200, 300, 400, 500$ について最小の MSE が得られる r と N を関数 f_i ($i = 1, 2, 3, 4$) についてそれぞれ求め、 $r = 0.14, 0.12, 0.06, 0.06$ 及び $N = 500, 500, 200, 400$ という値を得た。パラメータ

c_i と w_i はこう配法で求めたが、その学習係数は f_1 と f_2 については 0.05, f_3 と f_4 については 0.01 とした。これは f_3 と f_4 についての学習係数を 0.05, 0.04, 0.03, 0.02 とすると MSE はある大きな値からほとんど減少しなかったからである。SVR については SVM^{light} の手法 [5] を用い、不感度 $\epsilon = 0.001$ 及び関数 f_i ($i = 1, 2, 3, 4$) に対するガウスクERNEL $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/r^2)$ の幅 r をそれぞれ $r = 0.34, 0.34, 0.20, 0.20$ とすると次のすべてのパラメータ値の組合せの試行に対し最小の MSE を実現した: $\epsilon = 0.01, 0.001, 0$, $r = 0.04, 0.06, \dots, 0.50$ 及び訓練データ数 $t = i \times 10^3$ ($i = 1, 2, 5, 10, 50$)。なお CAN2 と公平な比較を行うには $t = 5 \times 10^7$ が必要であると考えられるが、 $t = 50 \times 10^3$ としたときでも 2 次元最適化の繰返しが十分に収束せず、非常に長い計算時間を要する場合があります、そのような場合は繰返し回数の上限を 10^6 として打ち切った結果を用いた。これと同様の現象は [5] にもあるように $\epsilon = 0$ の場合にも生じ同様の対策を行った。なお、以後、SVR の訓練データ数 t は他手法の学習回数 t と同一視し、生成されたサポートベクトル数 N は他手法のユニット数 N と同一視する。また以下の RBFN と SVR の結果は特に言及しない限り上述の試行で求めた最小の MSE を与えるパラメータ値を使っている。

5.2 ユニット数 N に対する MSE

実験結果としてまず図 3 にユニット数 N に対する MSE を示す。ここで SVR 以外は $N = 5, 50, 100, 200, 300, 400, 500$ に対する MSE を結んだ折れ線、SVR については訓練データ数 $t = i \times 10^3$ ($i = 1, 2, 5, 10, 50$) に対して生成されたサポートベクトル数 N に対する MSE を順に結んだ折れ線である。SVR の折れ線のほとんどは t が大きくなると N は大きく MSE は小さくなるので右肩下がりになるが、そうでない場合もある。これは大きな t では最適化の繰返しが十分に収束しない場合があった(前述)ことが原因であると考えられる。さて図 3 の CAN2 の MSE はすべての関数に対して N の増加に伴って単調に減少し、かつ CAN2-OR よりも小さいかほぼ等しい MSE を実現している。また関数 f_4 については SVR が最も小さな MSE を達成したがその他の関数については N が十分大きいと CAN2 (あるいは CAN2-OR) は他手法よりも小さい MSE を実現していることがわかる。関数 f_4 は CAN2 の漸近最適性の成立条件を満たさないで、その MSE があまり小さくならなかったことは

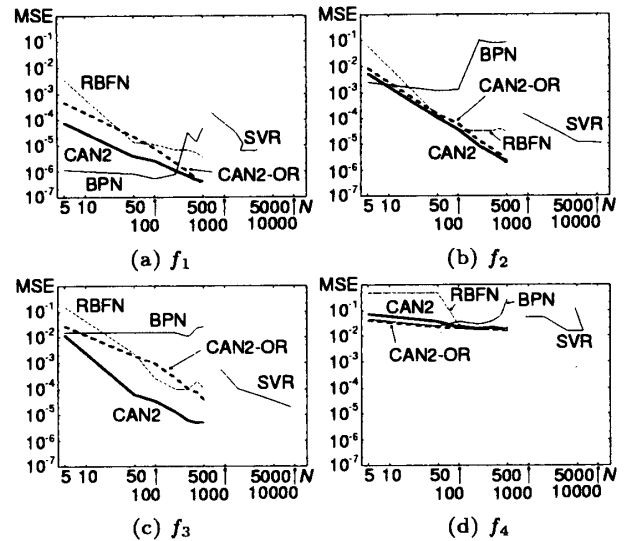


図 3 ユニット数 N に対する MSE
Fig. 3 MSE vs. N .

理解できるが、このほかに f_1 に対する $N = 500$ の場合も CAN2 の MSE は CAN2-OR よりもわずかに大きかった。これは CAN2 の漸近最適条件の逐次計算においては統計的誤差が避けられないことが本質的な原因であると考えられる。例えば雑音 d の影響を吸収するには十分な訓練データが必要であり、式 (22) で $T_{RI} = 200N$ としたが、この設定が小さかったことが挙げられる。これは学習回数 t に対する MSE を示す図 4 (詳しくは下述) の (a) を見ると CAN2 の MSE は $t = 10^5$ から既に CAN2-OR の MSE よりも大きくその時点までの CAN2 の再初期化が適当でなかったことを示していることから判断できる。 T_{RI} を非常に大きくすれば両者の差はなくなるが、あまり大きくしすぎると再初期化が機能なくなり、またこの MSE の差は他の結果と比べると無視できるほどの差であるので再設定して再実験することは行わなかった。更に逐次計算の性能を左右するパラメータ値として式 (14), (15) の τ_η を $800N$ 及び式 (21) の θ_α を 2 としたがこれらの影響の検討は今後の課題である。

5.3 学習回数 t に対する MSE

図 3 で各関数に対して最も良い結果を示した N の値のときの学習回数 t に対する MSE を図 4 に示す。この図 (a) と (d) の BPN に振動が見られるがこれは雑音 d の影響であり、 $d = 0$ とするとこの振動は消えた。更に同図 (a) については学習を $t = 5 \times 10^8$ まで行くと CAN2 よりも小さな MSE の値 6.2×10^{-8} まで減少したがこれは関数 f_1 が BPN で学習の容易なシグモ

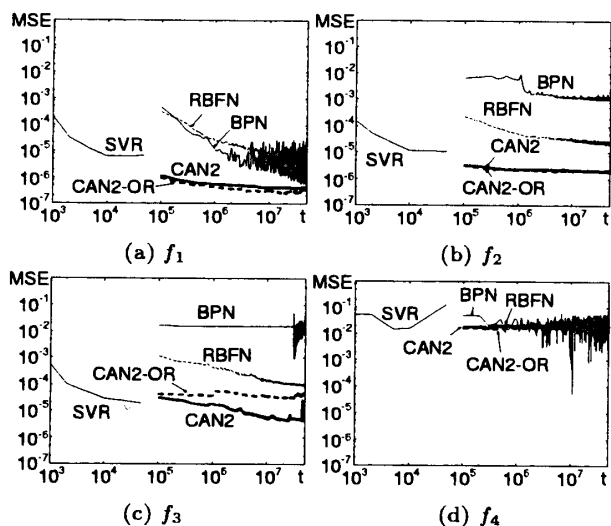


図4 学習回数 t に対する MSE
Fig. 4 MSE vs. t .

イド関数で定義されているからと考えられる。しかし雑音がある場合は CAN2 の MSE の値 4.1×10^{-7} よりも小さな MSE は実現できなかった。なお関数 f_2 は複数のガウス関数の和などで定義されており RBFN や SVR は BPN よりも良い結果を示したが N が大きな CAN2 や CAN2-OR には及ばなかった (図 3)。さて図 4 より CAN2 は学習回数 t が小さくても ($t = 10^5$ 程度でも) かなり小さな MSE を実現することがわかる。なお同図 (c) においては t の増加に対して CAN2 の MSE の振動が見られるが、この振動は再初期化によるものでありかつ観測雑音 d を加えない場合は起こらなかった。したがってこの振動は漸近最適条件の逐次計算が雑音 d をうまく吸収できずに不適切な再初期化が行われたから生じたといえる。

5.4 CAN2 の荷重ベクトル分布

CAN2 の N が 100 と 500 の場合の各関数学習後の荷重ベクトル w_i の分布を図 5 と図 6 に示す。図より関数の非線形性が強いところにより多くの w_i が集まり、与えられたユニット数で与えられた関数をより良く (区分的線形) 近似するための w_i の分布が形成されていると考えられる。これは量子化係数 C_i が非線形性の強さ $\gamma_i^{2/k}$ に比例するという予想 (3.1 参照) と、漸近最適条件式 (13) は各 w_i のポロノイ領域の体積 v_i が $(C_i p_i)^{1/(1+4/k)}$ に反比例することを意味する、ということを反映していると考えられる。なお図 6 (d) の関数 f_4 について不連続な十字の周囲に w_i が少ないのは、その領域で競合に勝つ荷重ベクトル w_c はこう配法の式 (A-5) により、一般に誤差が大

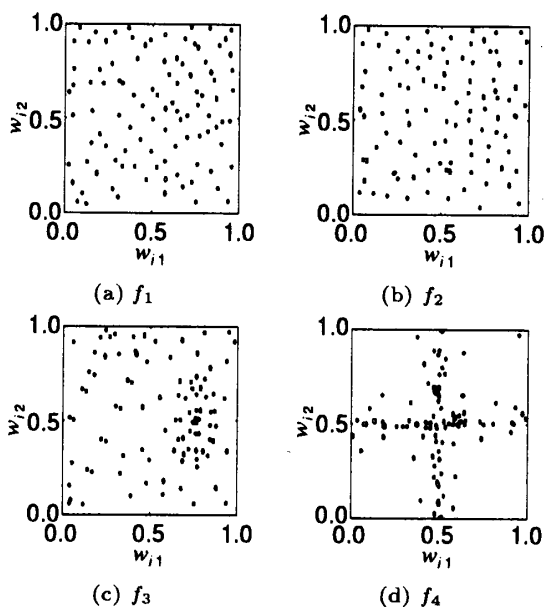


図5 CAN2 の w_i の分布 ($N = 100$)
Fig. 5 Distribution of w_i of CAN2 ($N = 100$).

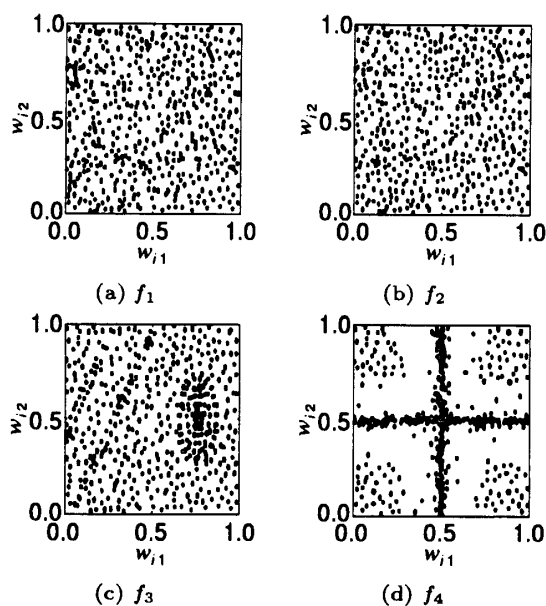


図6 CAN2 の w_i の分布 ($N = 500$)
Fig. 6 Distribution of w_i of CAN2 ($N = 500$).

きな不連続な十字領域の方向に移動するからである。

5.5 MSE データと漸近最適性

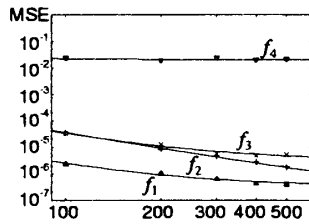
さて CAN2 が漸近最適条件を満たすならばその MSE は式 (12) の最下段の式に準じて N の増加に伴い減少すると考えられる。ただし式 (12) は雑音を加えた訓練データに対する MSE の期待値を表すが、本実験の MSE は雑音を加えた訓練データを学習したネットの雑音を加えないテストデータに対する MSE であ

| | <i>a</i> | <i>b</i> | <i>c</i> | <i>b_L</i> | <i>b_U</i> |
|-----------------------|----------------------|----------|----------------------|----------------------|----------------------|
| <i>f</i> ₁ | 5.9×10^{-3} | 1.7 | 2.1×10^{-7} | 0.4 | 3.0 |
| <i>f</i> ₂ | 1.1 | 2.3 | 1.2×10^{-6} | 2.0 | 2.5 |
| <i>f</i> ₃ | 9.9×10^{-2} | 1.7 | 2.4×10^{-6} | 0.9 | 2.5 |
| <i>f</i> ₄ | 6.8×10^4 | 3.7 | 2.0×10^{-2} | -49 | 58 |

(a)

| | <i>a</i> | <i>c</i> |
|-----------------------|----------------------|----------------------|
| <i>f</i> ₁ | 2.2×10^{-3} | 3.8×10^{-7} |
| <i>f</i> ₂ | 3.6×10^{-1} | 3.1×10^{-7} |
| <i>f</i> ₃ | 3.1×10^{-1} | 3.6×10^{-6} |
| <i>f</i> ₄ | 2.3×10 | 2.0×10^{-2} |

(b)



(c)

図 7 CAN2 の MSE の非線形最小 2 乗回帰結果
Fig. 7 Result of nonlinear least square regression of the MSE achieved by the CAN2.

ると同時にネットのすべての計算誤差を含むことを注意しておく。まず図 3 の CAN2 の MSE のうち *N* が十分大きいと考えられる *N* = 100, 200, ..., 500 について非線形最小 2 乗回帰により $MSE = aN^{-b} + c$ の係数 *a*, *b*, *c* を求めた結果を図 7(a) に示す。この図で区間 (*b_L*, *b_U*) は係数 *b* の 95% の信頼区間を表し、式 (12) の減少の度合 $4/k = 2$ を含んでいることがわかる。そこで再度、 $MSE = aN^{-2} + c$ として非線形最小 2 乗回帰を行うと同図 (b) が得られ、その回帰曲線 (同図 (c)) は実験データをよく再現していると判断できる。さて *f*₁ と *f*₂ に対して得られた同図 (b) の *c* の値は CAN2 による計算誤差のうち本論文の漸近最適化によっては小さくできない成分を表すといえる。なお雑音 *d* を加えない場合もこれと同程度の MSE しかなかったためこの *c* の値は雑音 *d* の逐次的統計処理に起因する誤差はほとんど含んでないと考えられる。一方、*f*₃ と *f*₄ に対する *c* の値は *f*₁ や *f*₂ に対する *c* の値のそれぞれ約 10 倍及び約 10⁵ 倍あるが、これらの値は漸近最適化では減少できない成分であると同時に、その大部分は各関数が漸近最適性の成立条件を満たさない程度を表していると考えられる。

5.6 入力分布の影響

一様分布でない入力分布の訓練データに対しても CAN2 がうまく学習できるかどうかを調べるために、次の三つの分布に従う訓練データを学習させ、前述の

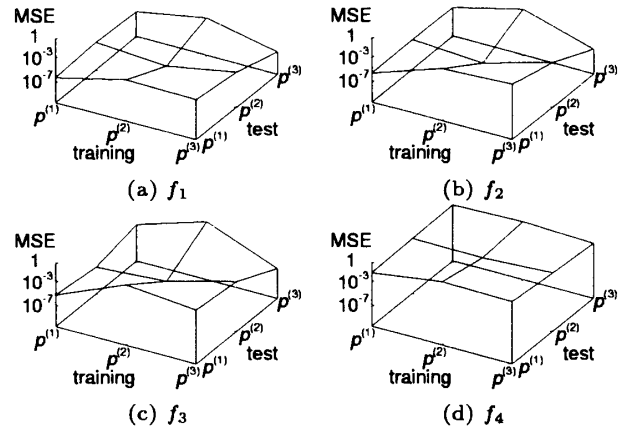


図 8 訓練分布とテスト分布に対する CAN2 の MSE
Fig. 8 MSE of the CAN2 for distributions of training and test data.

33 × 33 の格子上的テストデータに対する 2 乗誤差を各分布に基づいて重み付けした MSE を計算した：

$$p^{(1)}(\mathbf{x}) = \text{一定} \tag{23}$$

$$p^{(2)}(\mathbf{x}) = \exp\left(-\frac{(x_1 - \mu_1)^2 + (x_2 - 1/2)^2}{2\sigma^2}\right) \tag{24}$$

$$p^{(3)}(\mathbf{x}) = \exp\left(-\frac{(x_1 - \mu_2)^2 + (x_2 - 1/2)^2}{2\sigma^2}\right) \tag{25}$$

ここで $\mu_1 = 1/3$, $\mu_2 = 2/3$ 及び $\sigma = 0.1$ とし、各分布は単位正方形 $S = \{(x_1, x_2) \mid x_1, x_2 \in [0, 1]\}$ の外では 0 とし *S* 内で正規化した。ユニット数 *N* = 500 の場合の実験結果を図 8 に示す。図より訓練分布と学習分布が同じ場合に小さな MSE が実現されていることがわかる。ただし関数 *f*₃ のテスト分布が *p*⁽²⁾ のときは、訓練分布が *p*⁽²⁾ のときの MSE の値 4.5×10^{-6} よりも *p*⁽¹⁾ のときの MSE の値 8.0×10^{-7} の方が小さかった。これは訓練した確率密度 *p*⁽²⁾ の値が非常に小さい領域における誤差が影響していると考えられる。なぜならその領域の誤差は一様分布 *p*⁽¹⁾ で学習した方がより小さくなると考えられること、及び確率密度が小さい領域では連想行列 *M_i* の学習が十分行われているという漸近最適条件を導出した際の仮定が成立しなくなる可能性があるからである。この問題の解決及び CAN2 がどれくらいの分布の違いに適應できるのかを更に詳しく解析することは実際上重要であり、今後の課題である。

6. むすび

本論文では競合連想ネット CAN2 の漸近最適条件を導出し、逐次学習に組み入れるとともに、他手法との比較を含む数値実験によりその有効性を確認した。本論文の結論として以下の点は特に重要である。

(1) 本論文の漸近最適性は、いくつかの仮定を必要とするが、局所最適性でなく大域最適性を与える(式(12)参照)。

(2) BPN, RBFN, SVR などの手法ではより良い結果を与えるユニット数 N を求めるため多くの時間を費やすことがあるが、漸近最適性が成立するとその必要性が軽減する。

(3) 本論文の数値実験における RBFN と SVR のパラメータ値は各学習関数に対してそれぞれ数多くの試行を行いより良い結果を与える値を求める手法 [4], [5] を用いたが、CAN2 は同じパラメータ値で各関数を学習しかつ非常に良い近似性能を示した。

最後に、ガウス関数を用いる RBFN や SVR による関数近似はその局所性のため外挿能力はあまり期待できないが、CAN2 は区分的に線形な外挿を行うことができる。適応制御 [8] や降水量推定 [10] などの実際の応用事例においては入出力関数が特定できず外挿を含めた関数近似能力(あるいは汎化能力)が要求される場合が多いので、今後、このような能力についても検討する予定である。

謝辞 本研究の一部は文部科学省科研費(課題番号 12680389)の援助を受けました。ここに感謝致します。

文 献

- [1] S. Kurogi and S. Ren, "Competitive associative networks for function approximation and control of plants," Proc. NOLTA '97, pp.775-778, 1997.
- [2] N. Dyn, D. Levin, and S. Rippa, "Data dependent triangulations for piecewise linear interpolation," IMA Journal of Numerical Analysis, vol.10, pp.137-154, 1990.
- [3] T. Jaditz and L.A. Riddick, "Time-series near-neighbor regression," Studies in nonlinear dynamics and econometrics, vol.4, no.1, pp.35-44, 2000.
- [4] M.J.L. Orr, "Recent advances in radial basis function networks," Technical Report, Institute for Adaptive and Neural Computation Division for Informatics, Edinburgh University, 1999.
- [5] T. Joachims, "Making large-scale SVM learning practical," in Advances in Kernel Methods - Support Vector Learning, chapter 11, MIT Press, 1998.
- [6] S. Kurogi, T. Sakamoto, and S. Ren, "An analysis of competitive associative nets," Proc. ICONIP '98,

pp.683-686, 1998.

- [7] S. Kurogi, H. Hirata, T. Nishida, and T. Sakamoto, "Model-switching adaptive control using competitive associative nets," Proc. NOLTA '99, vol.2, pp.755-758, 1999.
- [8] S. Kurogi, T. Nishida, T. Sakamoto, K. Itoh, and M. Mimata, "A simplified competitive associative net and a model-switching predictive controller for temperature control of chemical solutions," Proc. ICONIP-2000, pp.791-796, 2000.
- [9] 黒木秀一, 西田 健, "複数のモデルの学習と切り替えを行う競合連想ネットを用いる適応予測制御," 計測自動制御学会論文集, vol.37, no.3, pp.203-212, 2001.
- [10] 黒木秀一, 藤 誠, 寺田精一, "競合連想ネットを用いる降水量推定," 2001 信学総大, SD-1, March 2001.
- [11] A. Gersho, "Asymptotically optimal block quantization," IEEE Trans. Inf. Theory, vol.IT-25, no.4, pp.373-380, 1979.
- [12] 西田 健, 黒木秀一, 佐伯知則, "再初期化法を用いた適応ベクトル量子化," 信学論 (D-II), vol.J84-D-II, no.7, pp.1503-1511, July 2001.
- [13] C. Chinrungrueng and C.H. Séquin, "Optimal adaptive k-means algorithm with dynamic adjustment of learning rate," IEEE Trans. Neural Networks, vol.6, no.1, pp.157-169, 1995.
- [14] D. Wettschereck and T. Dietterich, "Improving the performance of radial basis function networks by learning center locations," in Advances in neural information processing systems 4, pp.1133-1140, Morgan Kaufmann Publishers, San Mateo, CA, 1992.

付 録

1. M_c と w_c の更新処理

M_c を更新するための逐次的最小 2 乗法には厳密解を求める方法 [9] もあるが、本論文ではより簡単な RLS (Recursive Least Squares) 法 [8] を用いて、

$$M_c := M_c + \frac{(y - M_c \tilde{x}) \tilde{x}^T \Psi_c}{1 + \tilde{x}^T \Psi_c \tilde{x}}, \quad (\text{A}\cdot 1)$$

により M_c を更新する。ここで $\Psi_c \in \mathbb{R}^{(k+1) \times (k+1)}$ も逐次的に次式により更新する。

$$\Psi_c := \Psi_c - \frac{\Psi_c \tilde{x} \tilde{x}^T \Psi_c}{1 + \tilde{x}^T \Psi_c \tilde{x}}. \quad (\text{A}\cdot 2)$$

M_c と Ψ_c の初期値はそれぞれ $M_c = \mathbf{O}$ と $\Psi_c = a\mathbf{I}$ とする。ここで \mathbf{O} はゼロ行列、 \mathbf{I} は単位行列、及び a は小さな定数である ($a = 0.001$ とした)。

次に式 (5) のエネルギー E の w_c の変化に対するこう配は

$$\frac{\partial E}{\partial w_c} = \sum_{i \in A_c} \int_{W_{ci}} (e_c^2(\mathbf{x}) - e_i^2(\mathbf{x})) p(\mathbf{x})$$

$$\frac{\mathbf{x} - \mathbf{w}_c}{\|\mathbf{w}_i - \mathbf{w}_c\|} \cos \theta_{icx} d\mathbf{x} \quad (\text{A.3})$$

となる [9]. ここで A_c は領域 V_c の隣接領域 V_i の添字 i の集合であり, 境界 W_{ci} は V_c と V_i の境界である. また θ_{icx} は W_{ci} の法線ベクトルと $\mathbf{x} - \mathbf{w}_c$ との角度であるが, V_c の周りに十分多くの V_i がある場合は $\theta_{icx} \approx 0$ と考えられるので以下では $\cos \theta_{icx} = 1$ とする. このよう配の逆方向に \mathbf{w}_c を更新すれば E は減少するので, \mathbf{x} が幅 ϵ ($\epsilon = 0.1$ とした) の境界 W_{ci} に入るとき, すなわち

$$\frac{|(2\mathbf{x} - \mathbf{w}_i - \mathbf{w}_c)^T (\mathbf{w}_i - \mathbf{w}_c)|}{\|\mathbf{w}_i - \mathbf{w}_c\|^2} \leq \epsilon \quad (\text{A.4})$$

を満たすとき,

$$\mathbf{w}_c := \mathbf{w}_c + s \left(\frac{\alpha(e_i^2(\mathbf{x}) - e_c^2(\mathbf{x}))}{\|\mathbf{w}_i - \mathbf{w}_c\|} \right) (\mathbf{x} - \mathbf{w}_c) \quad (\text{A.5})$$

により \mathbf{w}_c を更新する. ここで \mathbf{w}_c の初期値はゼロベクトル, α はこのユニットの発火回数 n_c の増加とともに減少する関数 $\alpha = 1/(1 + n_c)$ とする. 更に

$$s(\xi) = \begin{cases} 1 & (\xi \geq 1) \\ \xi & (1 > \xi > -0.1) \\ -0.1 & (\xi \leq -0.1) \end{cases} \quad (\text{A.6})$$

により更新量を制限するが, これはこう配法においては更新量が大きすぎるとエネルギーを減少させることができなくなるからであり, 更に上式で ξ が負のときは \mathbf{w}_c を \mathbf{x} から遠ざける処理になるのでその制限をより厳しくした. また確率近似法によると式 (A.4) を満たす現在の入力 \mathbf{x} に対して式 (A.5) により \mathbf{w}_c を更新していくと, 統計的には式 (A.4) を満たすすべての \mathbf{x} に対して更新を施すことになる. これは各時刻の入力 \mathbf{x} に対してすべての隣接ベクトル \mathbf{w}_i ($i \in A_c$) を求める必要はなく, 式 (A.4) を満たす一つの隣接ベクトル \mathbf{w}_i を求めればよいことを意味する. 更に \mathbf{w}_c の周りに十分多くの荷重ベクトル \mathbf{w}_j があるときはそれらの中に隣接ベクトル $\mathbf{w}_i \in A_c$ が存在することになる. そこで入力 \mathbf{x} に j 番目に近い荷重ベクトル \mathbf{w}_i ($i = c_j$) について式 (A.4) を $j = 2$ から N_c まで順に調べはじめに式 (A.4) を満たす $\mathbf{w}_i = \mathbf{w}_{c_j}$ についてのみ式 (A.5) を適用した. 以上の \mathbf{w}_c の更新の定常解は一般には局所最適解に収束するが, 最小化すべきエネルギーは下に有界なので, (大域的) 最適解の十分近くに荷重が配置されれば最適解に収束する.

2. k 次元直方体の正規化 2 乗誤差

式 (10) を成立させる凸多面体の一つとして, $\mathbf{x} = \mathbf{w}_i$ を中心とし, 各辺の向きが式 (6) の \mathbf{A}_i の固有ベクトルに平行で各辺の長さが $2r_{ij}$ ($j = 1, 2, \dots, k$) になる k 次元直方体 R_i について検討する. この計算は $\mathbf{x} = \mathbf{P}_i \mathbf{z}_i + \mathbf{w}_i$ とおいて式 (7) の E_i を変形すると簡単になる.

$$E_i = p_i \int_{R_i} \left(b_{i0} - M_{i0} - \mathbf{m}_i^T \mathbf{w}_i + (\mathbf{b}_i - \mathbf{m}_i)^T \mathbf{P}_i \mathbf{z}_i + \mathbf{z}_i^T \mathbf{\Lambda}_i \mathbf{z}_i + d \right)^2 d\mathbf{z}_i \quad (\text{A.7})$$

ここで $b_{i0} = a_{i0} + (\mathbf{a}_i^T + \mathbf{w}_i^T \mathbf{A}_i) \mathbf{w}_i$, $\mathbf{b}_i = \mathbf{a}_i + 2\mathbf{A}_i \mathbf{w}_i$, 及び $\mathbf{m}_i = (M_{i1}, M_{i2}, \dots, M_{ik})^T$ とした. また行列 \mathbf{A}_i の固有値を λ_{ij} ($j = 1, 2, \dots, k$) とし, 正規化した固有ベクトルを \mathbf{p}_{ij} とし, 更に \mathbf{p}_{ij} を並べた行列を $\mathbf{P}_i = [\mathbf{p}_{i1}, \dots, \mathbf{p}_{ik}]$ とすると, \mathbf{A}_i が対称行列のとき \mathbf{P}_i は (正規) 直交行列になるという性質を用いた. また $\mathbf{\Lambda}_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ik}) = \mathbf{P}_i^T \mathbf{A}_i \mathbf{P}_i$ であり, この固有値 λ_{ij} は重複を含めたものである. なお以下では, 簡単のためこれらの固有値の中には 0 が含まれない場合を扱うが, 0 を含めた場合も同様の結論が得られる. さて $\partial E_i / \partial M_{ij} = 0$ より, E_i を最小化する M_{ij} として次式を得る.

$$M_{ij} = \begin{cases} b_{ij} & (j \neq 0), \\ b_{i0} + \frac{1}{3} \sum_{m=1}^k \lambda_{im} r_{im}^2 - \sum_{m=1}^k b_{im} w_{im} & (j = 0). \end{cases} \quad (\text{A.8})$$

これより最小の E_i は

$$E_i = \frac{p_i v_i}{45} \sum_{m=1}^k \lambda_{im}^2 r_{im}^4 + \sigma_d^2 p_i v_i \quad (\text{A.9})$$

となる. ここで $v_i = v(R_i) = \prod_{n=1}^k (2r_{in})$ は R_i の体積である. 更に相加平均 \geq 相乗平均の関係より

$$E_i \geq \frac{4p_i v_i k}{45} \left(\prod_{m=1}^k \lambda_{im}^2 r_{im}^4 \right)^{1/k} + \sigma_d^2 p_i v_i = \frac{2^{2-4/k} k p_i \gamma_i^{2/k} v_i^{1+4/k}}{45} + \sigma_d^2 p_i v_i \quad (\text{A.10})$$

となる. ここで $\gamma_i \triangleq \prod_{m=1}^k |\lambda_{im}|$ とした. この式

(A・10)の上段の不等式で等号が成立するのは $\lambda_{im}^2 r_{im}^4$ がすべての $m = 1, 2, \dots, k$ について一定になるときのみである。換言すると各辺の長さ r_{im} が $\sqrt{|\lambda_{im}|}$ に反比例する場合にのみ等号が成立し、そのとき体積 v_i を固定したときの E_i の最小値となる。更に上式より正規化2乗誤差は

$$I(R_i) = \frac{2^{2-4/k} k \gamma_i^{2/k}}{45} \quad (\text{A} \cdot 11)$$

となり、 $I(\xi R_i) = I(R_i)$ が成立する。この式はポリトープ H_i が k 次元直方体 R_i という制約下で E_i を最小にする正規化2乗誤差 $I(H_i)$ を与えるので、最適なポリトープ H_i^* の量子化係数 $C_i = I(H_i^*)$ は明らかにこれ以下になる。

3. k 次元楕円体の正規化2乗誤差

軸の向きを \mathbf{A}_i の固有ベクトル \mathbf{p}_{im} ($m = 1, 2, \dots, k$) 方向にとり軸長を $r_{im} = r/\sqrt{|\lambda_{im}|}$ とした k 次元楕円体、すなわち $\sum_m |\lambda_{im}| z_{im}^2 \leq r$ となる k 次元楕円体 B_i を考える。ここで r は正の定数である。まず対称性から $j \neq 0$ のときの $\partial E_i / \partial M_{ij} = 0$ の解 M_{ij} は k 次元直方体の場合と同じく $M_{ij} = b_{ij}$ が解となる。次に M_{i0} については次のような極座標系 $(\rho_i, \theta_{i1}, \dots, \theta_{i,k-1})$ に変換すると便利である。ここで $\rho_i > 0$, $0 \leq \theta_{im} \leq \pi$ ($m = 1, 2, \dots, k-2$), $0 \leq \theta_{i,k-1} \leq 2\pi$, 及び

$$z_{im} = \begin{cases} \rho_i |\lambda_{im}|^{-1/2} \cos \theta_{im} \prod_{l=1}^{m-1} \sin \theta_{il} & (1 \leq m \leq k-1) \\ \rho_i |\lambda_{im}|^{-1/2} \prod_{l=1}^{m-1} \sin \theta_{il} & (m = k) \end{cases} \quad (\text{A} \cdot 12)$$

である。直交座標系 (z_{i1}, \dots, z_{ik}) からの座標変換のためのヤコビアンは

$$|J| = \left| \frac{\partial(z_{i1}, \dots, z_{ik})}{\partial(\rho_i, \theta_{i1}, \dots, \theta_{i,k-1})} \right| = \rho_i^{k-1} \gamma_i^{-1/2} \left(\prod_{m=1}^{k-2} (\sin \theta_{im})^{k-m-1} \right) \quad (\text{A} \cdot 13)$$

となる。するとエネルギー E_i は

$$E_i = p_i \gamma_i^{-1/2} \int_0^{2\pi} \int_0^\pi \dots \int_0^\pi \int_0^r (b'_{i0} - M_{i0} + \rho_i^2 + d)^2 \rho_i^{k-1}$$

$$\cdot \left(\prod_{m=1}^{k-2} (\sin \theta_{im})^{k-m-1} \right) d\rho_i d\theta_{i1} \dots d\theta_{i,k-1} \quad (\text{A} \cdot 14)$$

となる。ここで $b'_{i0} = b_{i0} - \sum_{m=1}^k M_{im} w_{im}$ である。よって $\partial E_i / \partial M_{i0} = 0$ より、 $M_{i0} = b'_{i0} + kr^2 / (k+2)$ を得る。したがって最小の E_i は

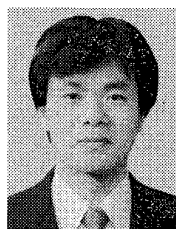
$$E_i = \frac{4p_i \gamma_i^{-1/2} k \pi^{k/2} r^{k+4}}{\Gamma(1+k/2)(k+2)^2(k+4)} + \sigma_d^2 p_i v_i \quad (\text{A} \cdot 15)$$

となる。ここで $\Gamma(n)$ はガンマ関数である。楕円体 B_i の体積は $v_i = \gamma_i^{-1/2} \pi^{k/2} r^k / \Gamma(1+k/2)$ であるから、正規化2乗誤差は

$$I(B_i) = \frac{4\pi^2 k \gamma_i^{2/k}}{\Gamma(1+k/2)^{4/k} (k+2)^2 (k+4)} \quad (\text{A} \cdot 16)$$

となり、 $I(\xi B_i) = I(B_i)$ が成立する。

(平成 14 年 4 月 24 日受付, 9 月 4 日再受付)



黒木 秀一 (正員)

昭 55 九工大・工・電気卒。昭 60 東工大大学院博士課程了。同年より九工大・制御・助手を経て平 3 同大・助教授。工博。主にニューラルネットの研究に従事。日本神経回路学会、計測自動制御学会等各会員。