ファジィ結合した複数の局所的ニューラルネットワークによる 大域的関数近似

非会員 森下 謙太郎* 正員 相吉 英太郎*

Global Function Approximation by Fuzzy-Connection of Plural Local Neural Networks

Kentaro Morishita*, Non-member, Eitaro Aiyoshi*, Member

This paper presents new model that is called a fuzzy connected neural network. The fuzzy connected neural network is expected to be better approximation and converge faster than the conventional neural network. When input-output teaching data with non-linearity are widely over a range, it is necessary to increase the hidden-layer's neurons and neural parameters because of the simple structure of neural networks. So We propose to divided the input data space into plural sub-area in order to approximate the divided teaching data with a neural network specially prepared on the each divided space. The plural neural networks are connected with fuzzy membership functions to construct the fuzzy connected neural network on the whole input space. Plural neural networks are called local neural networks. Because each local neural network dose not learns all teaching data, it is expected that the surface of the square error function is more monotonous than it in the conventional neural network, and the learning problem of each local neural network can be solved more easily. These points give fuzzy connected neural network advantages over the conventional neural network, which needs repeated trial and error to tune weights. Finally this paper shows the ability of the proposal model.

キーワード：ニューラルネットワーク，ファジィ推論，ラジアル基底関数ネットワーク，関数近似

Keywords: neural network, fuzzy inference, radial basis function network, function approximation

1. はじめに

大規模なシステムのモデリングでは、近似精度のよい局所的な近似モデルを複数構成してから、それらを互いに結合して、大域的にも精度のよい近似モデルを構成するのが合理的といえる。本論文では、このような概念を近似モデルの一つである階層型ニューラルネットワーク（以降 N.N. と略記する）に適用する。具体的には、比較的広い範囲に存在する多数の教師データを学習する場合に、入力データ空間を複数の部分領域に分割し、その部分領域ごとに局所的 N.N. でその領域の教師データを近似する。そしてこれらを入力データ空間全域に渡って互いに接続して精度のよい大域的 N.N. を構成することを考える。

階層型 N.N. は 2 つの結合層からなる N.N. で、前段の結合係数と中間層ニューロンの特性関数によって中間ニューロンの個数分の基底関数が入力データ空間上に生成される。

そして十分な個数の中間ニューロン数のもとで、任意の非線形関数を任意の精度で近似できることが文献 (1) で証明されている。しかし、中間層への入力に N.N. 全体への入力をアフィン変換したもので、さらにそれを中間ニューロンの特性関数である単調なシグモイド関数で変換することから、近似対象の非線形性が強い場合や教師データの数が多い場合には、その近似能力は必ずしも優れたものではない。このため、前述のような局所的 N.N. の接続により大域的 N.N. を構成することは、N.N. の近似能力の向上に有用と考えられる。

本論文では、これら複数の局所的 N.N. の結合方法として、ファジィ理論におけるメンバーシップ関数により、入力の値に依存する形で N.N. の結合係数を加重・加算することとし、入力データ空間全域に対応できるファジィ結合型 N.N. を大域的 N.N. として構成する。これにより各局所的 N.N. は入力データ空間全体に対応する必要がないため、それらの負担が軽減されて近似精度を良好に保つことができ、同時にメンバーシップ関数を適切に設定すれば、大域的な近似精度の向上も期待される。

一方、if-then ルールにファジィ理論を適用して得られる

* 慶應義塾大学大学院理工学研究科

〒223-8522 横浜市港北区日吉 3-14-1

Graduate School of Science and Technology, Keio University
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522

ファジィ推論には、ラジアル基底関数ネットワーク (以降 R.B.F.N. と記す) とよばれるものがある。これは中間層ニューロン個数分のラジアル基底関数の 1 次結合による関数近似と解釈され、このネットワークも十分な個数の中間ニューロン数のもとで、任意の非線形関数を任意の精度で近似できることが文献 (2) などで証明されている。また階層型 N.N. よりも高速に教師データを学習できることが、文献 (3) において示されている。

本論文で提案するファジィ結合型 N.N. は、メンバーシップ関数 (ラジアル基底関数) を用いるという点では、ファジィ推論 (ラジアル基底関数ネットワーク) と共通点がある。しかし、ファジィ推論は、結論部が定数である複数の if-then ルールをファジィ接続したものであることから、入力データの部分領域上で定義された定数関数を接続していることになる。したがって、局所的 N.N. が実現する非線形関数を接続するファジィ結合 N.N. の方が、非線形性のより強い教師データに対する高い近似能力と、N.N. 特有の汎化能力が期待される。

2. ファジィ結合ニューラルネットワークの構成

入力端数 L 、中間層ニューロン数 M 、出力端数 N からなる一般的な 2 結合層 N.N. を考える。入力を x 、対応する出力を z 、また入力層と中間層間の結合係数を $V = \{v_{ml}\}$ 、中間層と出力層間の結合係数を $W = \{w_{nm}\}$ 、閾値係数を $\theta = \{\theta_m\}$ とすると、入出力関係は

$$z_n = \sum_{m=1}^M w_{nm} f \left(\sum_{l=1}^L v_{ml} x_l + \theta_m \right) \dots\dots\dots (1)$$

またはベクトル行列表現で

$$z = W f (Vx + \theta) \dots\dots\dots (2)$$

と表すことができる。なお、関数 f はシグモイド関数

$$f(u) = \frac{1}{1 + \exp(-u)} \dots\dots\dots (3)$$

とする。ここで式 (1) の形より N.N. の入出力関係は、 L 次元実空間 R^L から N 次元実空間 R^N への非線形写像であり、第 n 出力成分 z_n は、 M 個の中間ニューロンによる基底関数

$$F_m(x) = f \left(\sum_{l=1}^L v_{ml} x_l + \theta_m \right) \dots\dots\dots (4)$$

$$m = 1, \dots, M$$

を係数 $w_{nm}, m = 1, \dots, M$ で 1 次結合したものとみなすことができる。したがって、 M 種類の係数 $(v_{m1}, \dots, v_{mL}), m = 1, \dots, M$ と $\theta_m, m = 1, \dots, M$ は、同一形状の関数 f から M 種類の互いに異なる L 次元実空間上の基底関数式 (4) を生成するパラメータとみなすことができる。このような基底関数を用いた任意の実数値連続関数 $\phi(x)$ の近似可能性については、たとえば文献 (1) により保証されている。また教師データが与えられた時のパ

ラメータ V, W, θ を求める手法として、総誤差関数の最小化を最急降下法により行う逆誤差伝搬法⁽⁴⁾が知られている。

一方、N.N. と同様な階層型ネットワークの形状をしたものに、R.B.F.N. がある。これはファジィ理論に関連して、放射状のメンバーシップ関数 μ

$$\mu(r, \delta) = \exp(-\delta^2 r^2) \dots\dots\dots (5)$$

を基底関数に用いたものであり、入力端数 L 、中間層数 M 、出力端数 N とすると入出力関係は

$$z_n = \sum_{m=1}^M w_{nm} \mu(\|x - v_m\|, \delta_m) \dots\dots\dots (6)$$

$$n = 1, \dots, N$$

で与えられる。このような R.B.F.N. では多変数入力 x が、それと代表点 v_m とのユークリッドノルムにより 1 次元化され、広がりに関するパラメータ $\delta = \{\delta_m\}$ と式 (5) により M 個の基底関数が構成される。そして N.N. 同様にこれら基底関数が 1 次結合することにより出力 z を得る。この R.B.F.N. の近似可能性については文献 (2) について保証されている。またパラメータである代表点 $V = \{v_{ml}\}$ 、広がり δ 、結合係数 $W = \{w_{nm}\}$ を求める手法は、Schwenker⁽⁵⁾ により与えられている。この手法は、総誤差関数の最小化を最急降下法により行う前段階として、代表点 V や広がり δ を入力 x により推定するため、学習効率が向上している点の特徴である。

しかし、N.N. の式 (1) や R.B.F.N. の式 (6) による関数近似の精度については、定量的な考察はほとんどない。ただ、N.N. に関しては入力 x のアフィン変換 $Vx + \theta$ と単調なシグモイド関数 f により基底関数が生成されるため、近似対象関数 $\phi(x)$ の非線形性が強い場合には、その近似精度は必ずしも優れたものではないことが予想される。一方 R.B.F.N. では入力変数が多く、かつそれらの変数成分間で大きく異なる特性を持つ関数を近似対象関数 $\phi(x)$ とする場合は、その精度が劣化すると考えられる。

そこで本論文では、N.N. における x のアフィン変換 $Vx + \theta$ の係数 V や θ を x の非線形関数化したシグモイド関数 f による基底

$$G_m(x) = f \left(\sum_{l=1}^L v_{ml}(x) x_l + \theta_m(x) \right) \dots\dots\dots (7)$$

$$m = 1, \dots, M$$

を考え、さらに後段の結合係数も $W(x)$ のように関数化することを考える。これにより基底関数生成用中間ニューロンの特性関数に単調な式 (3) を用いながらも、基底関数パラメータを入力 x に依存させ、基底関数を入力に応じて変化させることにより N.N. による関数近似精度を向上させることが期待される。

この $V(x), W(x), \theta(x)$ の具体的な関数形を、ファジィ理論の考え方をを用いて定義する。まず x 空間上に C 個の代表点 $\Gamma = (\gamma^1, \dots, \gamma^C), \gamma^c = (\gamma_1^c, \dots, \gamma_L^c)^T$ を用

意し、これに対応してパラメータ $\mathbf{v} = (V^1, \dots, V^C)$, $\mathbf{w} = (W^1, \dots, W^C)$, $\boldsymbol{\theta} = (\theta^1, \dots, \theta^C)$, $\boldsymbol{\delta} = (\delta^1, \dots, \delta^C)^T$ を用意する。そして、 $\mathbf{x}$ 空間の代表点 γ^c に近い程係数 V^c , W^c , θ^c の値に重みを置くメンバーシップ関数 $\mu(\|\mathbf{x} - \gamma^c\|, \delta^c)$ を考える。具体的には複数の代表点 $\gamma^c, c = 1, \dots, C$ に対応した正規化されたメンバーシップ関数値を重み係数として、代表点の個数分和をとって結合することにより $\mathbf{x}$ の関数

$$V(\mathbf{x}) = \sum_{c=1}^C \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c) V^c}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})} \dots (8)$$

$$W(\mathbf{x}) = \sum_{c=1}^C \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c) W^c}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})} \dots (9)$$

$$\theta(\mathbf{x}) = \sum_{c=1}^C \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c) \theta^c}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})} \dots (10)$$

を定義する。以下では、

$$g^c(\mathbf{x}; \Gamma, \boldsymbol{\delta}) = \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c)}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})} \dots (11)$$

とおき、メンバーシップ関数 $\mu(r, \delta)$ は R.B.F.N. と同様に

$$\mu(r, \delta) = \exp(-\delta^2 r^2) \dots (12)$$

を用いるものとする。

以上のファジィ結合を N.N. の構造から解釈すれば、入力 $\mathbf{x}$ の空間における代表点 γ^c を中心とした局所的な領域に対応して、 V^c , W^c , θ^c を係数とする N.N. を C 個 ($c = 1, \dots, C$) 構成しておき、代表点間の中間領域ほどそれらをファジィ化して互いに結合していることになる。前者の N.N. を局所的 N.N. (Local Neural Network: 以降 L.N.N. と略記する)、これらのファジィ結合によって構成される後者の N.N. をファジィ結合 N.N. (Fuzzy Connected Neural Network: 以降 F.C.N.N. と略記する) と呼ぶことにする。このとき関数 g^c は、入力が $\mathbf{x}$ のときの F.C.N.N. に対する第 c L.N.N. の寄与度、または結合強度を表すと解釈することができる。以上 F.C.N.N. の入出力関係をまとめると、次のように表される。

$$\begin{aligned} z_n &= \sum_{m=1}^M w_{nm}(\mathbf{x}) f\left(\sum_{l=1}^L v_{ml}(\mathbf{x}) x_l + \theta_m(\mathbf{x})\right) \\ &= \sum_{m=1}^M \left(\sum_{c=1}^C g^c(\mathbf{x}; \Gamma, \boldsymbol{\delta}) w_{nm}^c \right) \\ &\quad \times f\left(\sum_{l=1}^L \left(\sum_{c=1}^C g^c(\mathbf{x}; \Gamma, \boldsymbol{\delta}) v_{ml}^c \right) x_l \right. \\ &\quad \left. + \sum_{c=1}^C g^c(\mathbf{x}; \Gamma, \boldsymbol{\delta}) \theta_m^c \right) \\ &\quad n = 1, \dots, N \dots (13) \end{aligned}$$

ここで、F.C.N.N. と R.B.F.N. との関係も確認する。ま

ず式 (13) を

$$z_n = \sum_{m=1}^M \left(\sum_{c=1}^C g^c(\mathbf{x}; \Gamma, \boldsymbol{\delta}) w_{nm}^c \right) \times f\left(\sum_{l=1}^L v_{ml}(\mathbf{x}) x_l + \theta_m(\mathbf{x})\right) \dots (14)$$

$$= \sum_{c=1}^C \sum_{m=1}^M \left\{ w_{nm}^c f\left(\sum_{l=1}^L v_{ml}(\mathbf{x}) x_l + \theta_m(\mathbf{x})\right) \right\} \times g^c(\mathbf{x}; \Gamma, \boldsymbol{\delta}) \dots (15)$$

$$= \sum_{c=1}^C \sigma_n^c(\mathbf{x}) g^c(\mathbf{x}; \Gamma, \boldsymbol{\delta}) \dots (16)$$

と書きなおす。ただし

$$\sigma_n^c(\mathbf{x}) = \sum_{m=1}^M w_{nm}^c f\left(\sum_{l=1}^L v_{ml}(\mathbf{x}) x_l + \theta_m(\mathbf{x})\right) \dots (17)$$

である。式 (11) の関数 g^c はメンバーシップ関数 μ を正規化したものであることに注目して、式 (6) と式 (16) を比較し F.C.N.N. を R.B.F.N. に対応づけする。すると式 (6) の M を C とした C 個の中間層ニューロン数をもつ R.B.F.N. において、式 (6) の 1 次係数 w_{nm} を $\sigma_n^c(\mathbf{x})$ のように入力 $\mathbf{x}$ に応じて変化する関数にしたものと式 (16) はみることが出来る。すなわち、R.B.F.N. はファジィ推論の結論部に相当する w_{nm} の定数関数をファジィ結合したもののであるのに対して、F.C.N.N. はこの定数関数を入力 $\mathbf{x}$ に応じて式 (17) のように変化する非線形関数にしたものといえる。このことから F.C.N.N. の方が R.B.F.N. よりも入力 $\mathbf{x}$ に対する適応能力が高いものと期待される。

3. ファジィ結合ニューラルネットワークによる関数近似と学習アルゴリズム

N.N. による関数近似は、近似対象とする関数 ϕ 上の有限個のサンプル点を教師データとして行われる。このサンプル数を P とし、 p 番目の学習データの入出力量を $(\mathbf{x}^p, \mathbf{d}^p)$ とする。なお、 $\mathbf{d}^p = \phi(\mathbf{x}^p)$ は入力量 $\mathbf{x}^p$ に対する N.N. の出力所望値である。したがって、入力 $\mathbf{x}^p$ に対する出力量 $\mathbf{z}^p$ と $\mathbf{d}^p$ との誤差を、 $p = 1, \dots, P$ に対して最小自乗の意味で最小化する近似問題は

$$\min \frac{1}{2} \sum_{p=1}^P \sum_{n=1}^N (d_n^p - z_n^p)^2 \dots (18)$$

$$\begin{aligned} \text{where } z_n^p &= \sum_{m=1}^M \left(\sum_{c=1}^C g^c(\mathbf{x}^p; \Gamma, \boldsymbol{\delta}) w_{nm}^c \right) \\ &\quad \times f\left(\sum_{l=1}^L \left(\sum_{c=1}^C g^c(\mathbf{x}^p; \Gamma, \boldsymbol{\delta}) v_{ml}^c \right) x_l^p \right. \\ &\quad \left. + \sum_{c=1}^C g^c(\mathbf{x}^p; \Gamma, \boldsymbol{\delta}) \theta_m^c \right) \end{aligned}$$

$$c = 1, \dots, C, p = 1, \dots, P \dots (19)$$

$$g^c(\mathbf{x}; \mathbf{\Gamma}, \delta) = \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c)}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})}$$

$$c = 1, \dots, C \dots \dots \dots (20)$$

と定式化することができる。

ところで、この問題の未定係数は $\mathbf{V}, \mathbf{W}, \mathbf{\Theta}, \mathbf{\Gamma}, \delta$ であり、入力量 $\mathbf{x}$ が与えられたもとでは、式 (18) の総誤差関数はこれらの未定係数の関数となり、

$$E(\mathbf{V}, \mathbf{W}, \mathbf{\Theta}, \mathbf{\Gamma}, \delta)$$

$$= \frac{1}{2} \sum_{p=1}^P \sum_{n=1}^N (d_n^p - z_n^p)^2 \dots \dots \dots (21)$$

と置くことができる。すなわちこの近似問題では関数 E を最小にするような C 個の L.N.N. の係数 $\mathbf{V}, \mathbf{W}, \mathbf{\Theta}$ と、それらの寄与度を規定する関数 g^c のパラメータ $\mathbf{\Gamma}, \delta$ を同時に求める問題として

$$\min_{\mathbf{\Gamma}, \delta, \mathbf{V}, \mathbf{W}, \mathbf{\Theta}} E(\mathbf{V}, \mathbf{W}, \mathbf{\Theta}; \mathbf{\Gamma}, \delta) \dots \dots \dots (22)$$

と定式化される。

また、先見的知識から適切なパラメータ $\mathbf{\Gamma}, \delta$ が与えられているもとでは、近似問題は関数 E を C 個の L.N.N. の係数 $\mathbf{V}, \mathbf{W}, \mathbf{\Theta}$ で最小化する問題と定式化される。さらに、前者の決定は後者による最小自乗誤差の値を考慮しながら試行錯誤的に求めるとすると

$$\min_{\mathbf{\Gamma}, \delta} \Psi(\mathbf{\Gamma}, \delta) \dots \dots \dots (23)$$

where $\Psi(\mathbf{\Gamma}, \delta)$

$$= \min_{\mathbf{V}, \mathbf{W}, \mathbf{\Theta}} E(\mathbf{V}, \mathbf{W}, \mathbf{\Theta}; \mathbf{\Gamma}, \delta) \dots \dots \dots (24)$$

と 2 段階の問題として定式化される。なお関数 Ψ は総誤差関数 E の $\mathbf{V}, \mathbf{W}, \mathbf{\Theta}$ のみに関する最小値で、残りの未定パラメータ $\mathbf{\Gamma}, \delta$ の関数になる。ここで各学習法を区別するために式 (22) の問題を解く学習法を 1 段階学習法 (single stage learning method), 式 (23)~式 (24) の問題を解く学習法を 2 段階学習法 (bi-stage learning method) と呼ぶことにする。

この 1 段階学習法は全てのパラメータを一度に最小化するため適切な手法を用いれば最適解が求ることができる。しかし勾配による最急降下法を用いると、ファジィパラメータである $\mathbf{\Gamma}$ は入力量 $\mathbf{x}$ の定義域外の値が、 δ は広がりとして無意味な値が求まってしまうことが文献 (6) により知られている。そのため 1 段階学習法を用いて $\mathbf{\Gamma}$ や δ の有用な値を求めることは必ずしも容易でないと想像される。

一方、2 段階学習法では代表点 $\mathbf{\Gamma}$ や広がり δ を 1 回更新する過程の中に、 $\mathbf{V}, \mathbf{W}, \mathbf{\Theta}$ に関する最小化演算が入れ子構造になっている。そのためニューラルネットワークに関するパラメータと、ファジィ結合に関するパラメータを別々に最小化する点が特徴となる。ここでファジィパラメー

タがニューラルネットワークのパラメータと比べてその数が少ないことや、かつそれらの値の最適解が厳密に求まなくても、対応するニューラルネットワークの学習を厳密に行えば精度よく近似できると考えられる。このため、パラメータを $(\mathbf{\Gamma}, \delta)$ と $(\mathbf{V}, \mathbf{W}, \mathbf{\Theta})$ とに分け、これらのパラメータに関する最小化演算を 2 段階に分割することは、数値計算上メリットがあると予想される。

また、最小化パラメータを分離する方法として、両パラメータを交互に最小化する交互学習法 (interactive learning method) も考えることができる。2 段階学習法では、ファジィパラメータにおける 1 回の微小更新につき、ニューラルネットワークパラメータのみ最小化を収束するまで行うのに対して、この交互学習法では、ニューラルネットワークパラメータの収束値を固定したもとで、ファジィパラメータの最小化を収束するまで行うことになる。これにより代表点 $\mathbf{\Gamma}$, 広がり δ の更新量が大きくなるので、対応するニューラルネットワークパラメータがすぐに収束してしまう可能性が減り、2 段階学習法よりもより最適解を求められるようになると思われる。

以下においては、式 (21) の勾配成分を導出する。まず表記簡略化のため $v_{ml}(\mathbf{x}) = v_{ml}$, $w_{nm}(\mathbf{x}) = w_{nm}$, $\theta_m(\mathbf{x}) = \theta_m$, $u_m^p = \sum_{l=1}^L v_{ml} x_l^p + \theta_m$, $y_m^p = f(u_m^p)$, $e_n^p = (d_n^p - z_n^p)$ として、誤差関数

$$E^p(\mathbf{V}, \mathbf{W}, \mathbf{\Theta}, \mathbf{\Gamma}, \delta) = \frac{1}{2} \sum_{n=1}^N (d_n^p - z_n^p)^2 \dots \dots (25)$$

を w_{nm}^c , v_{ml}^c および θ_m^c で偏微分すると

$$\frac{\partial E^p}{\partial w_{nm}^c} = \frac{\partial E^p}{\partial z_n^p} \frac{\partial z_n^p}{\partial w_{nm}^c} \frac{\partial w_{nm}^c}{\partial w_{nm}^c}$$

$$= - (d_n^p - z_n^p) y_m^p \frac{\partial w_{nm}^c}{\partial w_{nm}^c}$$

$$= - (d_n^p - z_n^p) y_m^p \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c)}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})}$$

$$= - e_n^p y_m^p \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c)}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})} \dots (26)$$

$$\frac{\partial E^p}{\partial v_{ml}^c} = \sum_{n=1}^N \left\{ \frac{\partial E^p}{\partial z_n^p} \frac{\partial z_n^p}{\partial y_m^p} \right\} \frac{\partial y_m^p}{\partial u_m^p} \frac{\partial u_m^p}{\partial v_{ml}^c} \frac{\partial v_{ml}^c}{\partial v_{ml}^c}$$

$$= - \left\{ \sum_{n=1}^N e_n^p w_{nm} \right\} y_m^p (1 - y_m^p) x_l^p$$

$$\times \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c)}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})} \dots \dots \dots (27)$$

$$\frac{\partial E^p}{\partial \theta_m^c} = - \left\{ \sum_{n=1}^N e_n^p w_{nm} \right\} y_m^p (1 - y_m^p)$$

$$\times \frac{\mu(\|\mathbf{x} - \gamma^c\|, \delta^c)}{\sum_{c'=1}^C \mu(\|\mathbf{x} - \gamma^{c'}\|, \delta^{c'})} \dots \dots \dots (28)$$

となる。

また、代表点 γ_l^c , 結合強度係数 δ^c についての偏微分を求めると、 $*$ = γ_l^c または $*$ = δ^c とし、 $g^c(\mathbf{x}^p; \mathbf{\Gamma}, \delta)$ の引

数 (Γ, δ) を省略すると

$$\begin{aligned} \frac{\partial E^p}{\partial \star} = & \sum_{n=1}^N \left[\frac{\partial E^p}{\partial z_n^p} \sum_{m=1}^M \left\{ \sum_{c'=1}^C \left(\frac{\partial g^{c'}(\mathbf{x}^p)}{\partial \star} w_{nm}^{c'} \right) y_m^p \right. \right. \\ & + w_{nm} \frac{\partial y_m^p}{\partial \left(\sum_{l=1}^L v_{ml} x_l^p \right)} \sum_{l=1}^L \left(\sum_{c'=1}^C \left(\frac{\partial g^{c'}(\mathbf{x}^p)}{\partial \star} v_{ml}^{c'} \right) \right. \\ & \left. \left. \times x_l^p + \sum_{c'=1}^C \left(\frac{\partial g^{c'}(\mathbf{x})}{\partial \star} \theta_m^{c'} \right) \right\} \right] \dots \dots \dots (29) \end{aligned}$$

となり、さらに

$$\begin{aligned} \frac{\partial g^{c'}(\mathbf{x}^p)}{\partial \star} = & \left\{ \begin{aligned} & \frac{\partial \mu(\|\mathbf{x}^p - \gamma^{c'=c}\|, \delta^{c'=c})}{\partial \star} \\ & \times \sum_{c''=1}^C \mu(\|\mathbf{x}^p - \gamma^{c''}\|, \delta^{c''}) \\ & - \mu(\|\mathbf{x}^p - \gamma^{c'=c}\|, \delta^{c'=c}) \\ & \times \frac{\partial \mu(\|\mathbf{x}^p - \gamma^{c''}\|, \delta^{c''})}{\partial \star} \end{aligned} \right\} \\ = & \left\{ \begin{aligned} & \div \left\{ \sum_{c''=1}^C \mu(\|\mathbf{x}^p - \gamma^{c''}\|, \delta^{c''}) \right\}^2, \\ & \text{if } c = c' \\ & - \mu(\|\mathbf{x}^p - \gamma^{c'=c}\|, \delta^{c'=c}) \\ & \frac{\partial \mu(\|\mathbf{x}^p - \gamma^{c''}\|, \delta^{c''})}{\partial \star} \\ & \div \left\{ \sum_{c''=1}^C \mu(\|\mathbf{x}^p - \gamma^{c''}\|, \delta^{c''}) \right\}^2, \\ & \text{if } c \neq c' \end{aligned} \right\}, \end{aligned}$$

および

$$\begin{aligned} \frac{\partial \mu(\|\mathbf{x}^p - \gamma^c\|, \delta^c)}{\partial \gamma_l^c} = & \exp(-\delta^{c2} \|\mathbf{x}^p - \gamma^c\|^2) \\ & \times (-2\delta^{c2} \|\mathbf{x}^p - \gamma^c\|^2) \frac{\partial \|\mathbf{x}^p - \gamma^c\|}{\partial \gamma_l^c} \\ = & 2\delta^{c2} (x_l^p - \gamma_l^c) \exp(-\delta^{c2} \|\mathbf{x}^p - \gamma^c\|^2) \cdot (30) \\ \frac{\partial \mu(\|\mathbf{x}^p - \gamma^c\|, \delta^c)}{\partial \delta^c} = & -2\delta^c \|\mathbf{x}^p - \gamma^c\| \\ & \times \exp(-\delta^{c2} \|\mathbf{x}^p - \gamma^c\|^2) \dots \dots \dots (31) \end{aligned}$$

を用いることにより、 $\frac{\partial E^p}{\partial \gamma_l^c}$ および $\frac{\partial E^p}{\partial \delta^c}$ を求めることができる。

4. 計算機シミュレーション

〈4・1〉 ファジィパラメータ固定学習 まず、ファジィ

パラメータ Γ, δ を既知として固定した下で総誤差関数を最小化するファジィパラメータ固定学習問題

$$\min_{\mathbf{V}, \mathbf{W}, \Theta} E(\mathbf{V}, \mathbf{W}, \Theta; \Gamma, \delta) \dots \dots \dots (32)$$

を解いた。そしてパラメータ Γ, δ 変えていくつシミュレーションを行うことにより、ファジィパラメータの依存特性を調べた。近似対象関数 $\phi(x)$ としては、関数

$$\phi(x) = \frac{\sin(2\pi x) + 1}{2} \quad 0 \leq x \leq 1 \dots \dots \dots (33)$$

上の区間 $[0, 1]$ における 21 個の点を学習データとして用いた。これを示したのが図 1 の○印である。

F.C.N.N. の構造は、 $L = N = 1$ で、 M と C については $(M, C) = (2, 5)$ および $(4, 2)$ の 2 種類とし、 $\mathbf{V}, \mathbf{W}, \Theta$ の初期値には区間 $[-3, 3]$ の一様乱数を与えた。

そして、代表点 Γ の依存性検証シミュレーションでは、広がり δ は一定とし、いずれの場合もすべての L.N.N. に共通で $\delta^c = 0.1, c = 1, \dots, C$ とした。また $L \times C$ 行列パラメータ Γ は、 $(M, C) = (2, 5)$ のとき、 $(0, 0.25, 0.5, 0.75, 1), (0.005, 0.255, 0.5, 0.745, 0.995), \dots, (0.49, 0.495, 0.5, 0.555, 0.55)$ の 50 通り、 $(M, C) = (4, 2)$ のとき、 $(0, 1), (0.005, 0.995), \dots, (0.49, 0.55)$ の 50 通りを与えて、それぞれについて式 (32) を最急降下法にもとづく学習法で解いた。これら代表点 Γ のパラメータは、 x 軸上の区間 $[0, 1]$ において、メンバーシップ関数をその区間全体に均一に配置させる場合から、0.5 付近に近寄らせて配置する場合まで、50 通りの配置のしかたを表している。

一方、広がり δ の依存性検証シミュレーションでは、 $\Gamma = (0, 0.25, 0.5, 0.75, 1)$ または $(0, 1)$ と一定とし、各 L.N.N. に共通でパラメータ δ を、 $0.01 \leq \delta^c \leq 63.0957, c = 1, \dots, C$ の範囲で 20 通り与えて、それぞれについて式 (32) を最急降下法にもとづく学習法で解いた。

これら総誤差 E の最小値と Γ や δ の関係を示したのがそれぞれ図 2 および図 3 である。なお、図 2 の横軸 γ は、上記の 50 通りの横ベクトル Γ の第 1 成分を代表させて表示した。また図 2、図 3 から、 $(M, C) = (2, 5)$ では、 $\Gamma = (0, 0.25, 0.5, 0.75, 1)$ 、 $\delta = 6.31$ のとき総誤差が最も小さくなっている。そこでこの時の近似関数の形を図 1 の×印で示す。また比較検討用に F.C.N.N. と同数のパラメータを有する N.N. に同様の区間 $[-3, 3]$ の一様乱数を与え、逆誤差伝搬法により学習を行った結果を+印で示す。

まず、総誤差関数の最小値の対 Γ 依存性を示す図 2 からは、パラメータ Γ の第 1 成分が 0 のとき、すなわち入力区間 $[0, 1]$ 上に等間隔にメンバーシップ関数の代表点をとるとき、総誤差関数の値が最小になっていることがわかる。これは式 (33) の近似では等間隔に教師データが取られているので、L.N.N. を入力空間 $[0, 1]$ 全体にわたって配置させる方が、総誤差関数の最小値をより小さくすることに効果があったためと考えられる。特に $(M, C) = (4, 2)$ の場合

図1 学習データと出力 (sin 関数)

Fig. 1. learning data and output data (sin).

図2 Γ による総誤差 E の変化

Fig. 2. the sum square error function E and the centers Γ .

図3 δ による総誤差 E の変化

Fig. 3. the sum square error function E and the widths δ .

より $(M, C) = (2, 5)$ の方が、上記のことが顕著に表れている。このことは中間層ニューロン数 M を増やして1個あたりの N.N. 近似能力を高めるより、中間層ニューロン数 M が少なくても L.N.N. の数 C を増やした方が効果があることがわかる。

一方、総誤差関数の最小値の対 δ 依存性を示す図3からは、メンバーシップ関数の広がり性を示すパラメータの値を区間 $[1, 10]$ の適切な値に設定すれば、総誤差関数の最小値をより小さくすることができることがわかる。パラメータ δ の値はこれら適切な値より大き過ぎても、複数の L.N.N. が分担する領域に重なりが生じて近似効果がなくなってしまう。逆に小さ過ぎたとしても、異なる L.N.N. 間の中間領域がカバーできないため誤差が生じるものと思われる。総誤差の最小値を与える δ は、 $C = 2$ のときは約 3.98, $C = 5$ のときは約 6.31 であることから、L.N.N. の個数が少ないとき（すなわち C が小さいとき）には、メンバーシップ関数の広がり性を大きくするために、式(5)における δ の値を小さくし、 C が大きい時はその逆になしなければならないことがわかる。

また図1を見ると、同様の条件下で学習を行った場合、F.C.N.N. の方が N.N. よりも精度よく近似ができている。以上のことから、パラメータである代表点 Γ や広がり δ の値を適切に設定した F.C.N.N. は、通常の N.N. と比較して近似精度を上げることができることがわかった。またこれらのパラメータは、近似対象の関数形や入力に依存するものの、入力領域全体にわたって等間隔に L.N.N. を配置し、広がりもそれらがうまく領域分担できるよう設定すれば良いことがわかる。

〈4・2〉 全パラメータ学習 つぎに1段階学習法、2段階学習法、交互学習法を用いて全パラメータを学習させるシミュレーションを行い、各学習法の比較結果を示す。

近似対象関数 $\phi(x)$ としては、図4のようなこぎり歯状関数

$$\phi(x) = \begin{cases} \sin\left(\frac{11}{8}\pi x\right) & 0 \leq x < \frac{9}{22} \\ \sin\left(\pi\left\{0.5 + \frac{11}{3}\left(x - \frac{8}{22}\right)\right\}\right) & \frac{9}{22} \leq x < 0.5 \\ \sin\left(\frac{11}{8}\pi(x - 0.5)\right) & 0.5 \leq x < \frac{20}{22} \\ \sin\left(\pi\left\{0.5 + \frac{11}{3}\left(x - \frac{19}{22}\right)\right\}\right) & \frac{20}{22} \leq x \leq 1 \end{cases}$$

の区間 $[0, 1]$ 上の23個のサンプル点を用いた。このように近似対象関数を関数の特性が断片的に異なるものとしたのは、学習データ空間の分割とそれに対する L.N.N. による近似およびその結合効果を確認するためである。

F.C.N.N. は、 $(M, C) = (2, 2), (2, 3), (2, 4), (3, 2),$

図4 学習データと出力 (のこぎり波状関数)

Fig. 4. learning data and output data (saw).

(4,2) の6種類の構造をテストした。これらの構造のF.C.N.N. に対して $\mathbf{V}$, $\mathbf{W}$, $\mathbf{\Theta}$ の初期値には、区間 $[-3, 3]$ の一様乱数を選んだ。またメンバーシップ関数の広がりを出すパラメータ δ の初期値には、いずれのF.C.N.N. を構成するすべてのL.N.N. に共通の値1を与え、メンバーシップ関数の代表点を与えるパラメータ行列 $\mathbf{I}$ の初期値には、その代表点が入力空間 $[0, 1]$ 全体にわたって等間隔になるよう設定した。これらの初期値のもとで、1段階学習法、交互学習法、2段階学習法のシミュレーションを10回ずつ行った。なお、各シミュレーションともパラメータ $\mathbf{V}$, $\mathbf{W}$, $\mathbf{\Theta}$ に関しては、初期値を10通り与えて試行した。これらの結果を表1に示す。表中の結合係数数とはF.C.N.N. の結合係数の総数、つまり $\mathbf{V}$, $\mathbf{W}$, $\mathbf{\Theta}$ の成分総数、パラメータ数はこの結合係数数にファジィパラメータ $\mathbf{I}$, δ の総数を加えたものである。また変更回数とは、交互学習法の場合 $(\mathbf{I}, \delta)$ に関する最小化段階と $(\mathbf{V}, \mathbf{W}, \mathbf{\Theta})$ に関する最小化段階を切り替えた回数、2段階学習法の場合は問題(23)の $(\mathbf{I}, \delta)$ の更新回数である。そして10回の試行の中で、これらの回数の最小回数-最大回数-平均回数を示している。さらに総誤差も10回の試行における、学習計算終了時の総誤差の最良値、最悪値、平均値を示している。また、比較用に同様の条件で、通常のN.N. とR.B.F.N. のシミュレーション結果も表2、表3に示す。そして図4にはパラメータ数が32となるF.C.N.N. の各学習法による最良値となったときの学習結果、および、比較用としてパラメータ数が33となるN.N. とR.B.F.N. の最良値となった学習結果を示す。

上記の結果より、1段階学習法、交互学習法、2段階学習法を比較すると、総誤差の値は1段階学習法と交互学習法は優劣つけがたく、2段階学習法はそれらより劣る傾向があることがわかる。交互学習法に比べて1段階学習法が劣る場合があるのは、種類が互いに異なるパラメータ $\mathbf{V}$, $\mathbf{W}$, $\mathbf{\Theta}$ とパラメータ $\mathbf{I}$, δ を同時に最適化するため、変数空間の次元が大きく、しかもそこで定義された誤差関数の形状も複雑で、局所的最適解も多くそれに陥るためと考

表1 F.C.N.N. の学習結果

Table 1. learning result (F.C.N.N.).

M	C	*1	*2	*3	*4	total value of error functions		
						bset value	worst value	mean value
2	2	12	16	single stage	-	0.3582	2.4605	1.4689
2	2	12	16	inter-active	2-66-15	0.4970	3.8337	2.0700
2	2	12	16	bi-stage	2-2-2	1.8823	2.3773	2.2309
2	3	18	24	single stage	-	0.1937	1.7581	0.7142
2	3	18	24	inter-active	1-114-19	0.2908	2.2685	1.2822
2	3	18	24	bi-stage	1-32-5	1.8908	2.3048	2.1431
2	4	24	32	single stage	-	0.0413	0.3220	0.2137
2	4	24	32	inter-active	1-50-17	0.2238	2.0717	1.4006
2	4	24	32	bi-stage	1-2-2	1.9356	2.3102	2.1255
3	2	18	22	single stage	-	0.1507	1.9558	1.0732
3	2	18	22	inter-active	1-28-9	0.3137	2.3196	1.0515
3	2	18	22	bi-stage	1-4-2	0.3163	2.3255	1.7309
4	2	24	28	single stage	-	0.3156	2.3090	1.6464
4	2	24	28	inter-active	1-83-13	0.2117	2.3364	1.1455
4	2	24	28	bi-stage	1-4-2	0.2117	2.3428	1.3985

*1:number of connective coefficients

*2:number of parameters

*3:learning method

*4:number of updating

表2 N.N. の学習結果

Table 2. learning result (N.N.).

M	number of connective coefficients	number of parameters	total value of error functions		
			best value	worst value	mean value
5	15	15	2.1576	2.2744	2.2214
6	18	18	2.1653	2.2752	2.2187
7	21	21	2.1003	2.3032	2.2364
8	24	24	2.1929	2.2702	2.2309
9	27	27	2.0859	2.2689	2.2273
10	30	30	1.9517	2.2847	2.1981
11	33	33	2.1931	2.2726	2.2178

表3 R.B.F.N. の学習結果

Table 3. learning result (R.B.F.N.)

M	number of connective coefficients	number of parameters	total value of error functions		
			best value	worst value	mean value
4	4	12	0.2540	11.0000	4.3267
5	5	15	0.1363	11.0003	4.4229
6	6	18	0.2368	11.0000	2.3452
7	7	21	0.2608	11.0000	3.7656
8	8	24	0.2308	11.0000	3.0253
9	9	27	0.2503	11.0000	2.5758
10	10	30	0.2212	11.0000	2.2955
11	11	33	0.2362	11.0000	2.9564

えられる。また、2段階学習法や交互学習法では、互いに性質の異なるパラメータの勾配計算を毎回の更新ごとに行わないので、計算時間が短縮される傾向があり、近似精度を要求しない場合には有効な学習法であるといえる。

5. おわりに

結合方法としてファジィ理論を用い、局所的モデルとしてN.N.を用いて大域的モデルを構成する方法を提案した。これは教師データ数が比較的多い学習を行わなければならない大規模なシステムのモデリングにおいて、近似精度の良い局所的な近似モデルを互いに結合して、大域的にも精度の良い近似モデルを構成するためである。この方法により、入力空間を複数の領域に分割し、個々の領域を分担するN.N.をより小さくすることにより、より少ない結合パラメータのN.N.でもある程度の近似精度を実現し、さらに学習効率をあげることができると期待される。また、この提案モデルに対する学習法を3種類提案し、計算機シミュレーションにより比較検討を行った。

近似精度の点では、メンバーシップ関数のパラメータ値を適切に設定すれば、総誤差関数の最小値を、通常のN.N.やR.B.F.N.よりも小さくすることができ、中でも1段階学習法が優れていることが確認された。一方、計算効率の点では、交互学習法が有効であることが示された。

R.B.F.N.と同様の役割を有するパラメータである入力値の代表点 Γ や広がり δ の学習法として、R.B.F.N.における試行探索法をこのF.C.N.N.に対しても用いることにより、一層近似精度が向上し、計算時間が短縮すると期待される。

(平成14年10月7日受付)

- (1) K. Funahashi: "On the approximation realization of continuous mappings by neural networks", *Neural Networks*, Vol.2, P.183-192 (1989)
- (2) E. Hartman, J. Keeler, and J. Kowalski: "Layered neural networks with gaussian hidden units as universal approximation", *Neural Computation*, 2, PP.210-215 (1990)
- (3) J. Moody, C. J. Darken: "Fast Learning in Networks of Locally-Tuned Processing Units", *Neural Computations*, 1, P.281-294 (1989)
- (4) D. E. Rumelhart, G. E. Hinton, and R. J. Williams: "Learning Representation by Back-Propagation of Errors", *Nature*, 323-9, P.533-536 (1986)
- (5) F. Schwenker, H. A. Kestler, and G. Palm: "Three learning phases for radial-basis-function networks", *Neural Networks*, Vol.14, No.4-5, p439-58, (2001-5)
- (6) D. Michie, D. J. Spiegelhalter, and G. C. Taylor: "Machine Learning, Neural and Statistical Classification", New York, Ellis Horwood (1994)

森 下 謙太郎 (非会員) 2000年慶應義塾大学理工学部物理情報工学科卒業。同年同大学大学院理工学研究科基礎理工学専攻前期博士課程進学。在学中は新しいニューラルネットワークや学習アルゴリズムの開発に関する研究に従事。

相 吉 英太郎 (正員) 1975年慶應義塾大学工学部計測工学科卒業。1980年同大学院工学研究科博士課程修了。専任講師、同理工学部助教授を経て、1996年より同教授。最適化の理論と計算方法に関する基礎研究に従事。工学博士。