

## 進化的多目的最適化法によるRBFネットワークのアンサンブル学習

近藤 伸彦\* 畠中 利治\* 魚崎 勝司\*\*

\*大阪大学大学院情報科学研究科情報数理学専攻

\*\*福井工業大学工学部経営情報学科

RBFネットワークの構造設計をモデルの複雑さに関する多目的最適化問題ととらえてこれを解くことで、多目的の意味で優れたRBFネットワークの集合を求めることができる。こうして求められた多様な構造のRBFネットワーク集合からアンサンブルを構築すれば、汎化能力に優れたモデルを構成できると考えられる。本研究では、進化的多目的最適化によるRBFネットワークの構築法によるアンサンブルを取り上げ、そのパターン分類への適用を検討した。アンサンブルを構成するメンバーの選択法と各メンバーの出力の結合法の組合せに対して、パターン分類のベンチマークテストデータを用いて数値実験を行った結果、従来のニューラルネットワークアンサンブルの手法と同等以上の能力を持つことが示された。

### The Ensemble Learning of RBF Networks by Evolutionary Multi-objective Optimization Approach

Nobuhiko Kondo\*, Toshiharu Hatanaka\* and Katsuji Uosaki\*\*

\*Department of Information and Physical Sciences, Osaka University

\*\*Department of Management and Information Sciences, Fukui University of Technology

The structure determination problem of RBF network can be considered as multi-objective optimization problem about model complexity. A set of RBF networks which is multi-objectively optimized can be obtained by solving the above problem. In this paper the construction of RBF networks by evolutionary multi-objective optimization method and its ensemble are considered, and it is applied to the pattern classification problem. In this paper, several combinations of ensemble member selection methods and output combination methods are considered. Experimental study on the benchmark problem of the pattern classification is carried out, then it is illustrated that the RBF network ensemble has comparable performance to the other ensemble methods.

#### 1 はじめに

ニューラルネットワークの一種である動径基底関数 (Radial Basis Function: 以下RBF) ネットワークは、優れた非線形写像能力と学習機能を有する数理モデルであり、関数近似能力や汎化能力についての理論からシステム同定やパターン認識などへの応用まで、多くの研究が行われてきた。RBFネットワークの性能は、RBFネットワークがいくつのニューロンから成りどのようにニューロンがリンクしているかというRBFネットワークの構造によって変化する。これは構造の複雑さに関するジレンマに起因している [1]。モデル構造をどのように決定するかという問題は実用上の観点からも重要であり、これまでも多くの研究がなされている。

一方で、近年、アンサンブル学習がその興味深い性質から機械学習やニューラルネットワークの分野において盛んに研究されている。アンサンブル学習は、複数の学習機械の出力を組み合わせることにより高い汎化能力を持たせることや複雑な問題を解くことを目指した学習法である。RBFネットワークのアンサンブルにおいては、個々のRBFネットワークの構築法と出力の組み合わせかたが問題となる [2]。

本研究では、多目的最適化の観点から得られたRBFネットワーク集合からアンサンブルを構築することを考える。まず、多目的最適化手法として優れた進化的計算によって、モデルの精度や構造の複雑さなどの評価にもとづいて学習されたRBFネットワークの集合を得る。こうして得られたRBFネットワーク集合はさまざまな構造を持ち、その多様性のためにアンサンブルを構成するメンバーとし

て適していると考えられるため、この集合から RBF ネットワークアンサンブルを構築する。この RBF ネットワークアンサンブルをパターン分類問題に適用し、数値シミュレーションによってそのふるまいを検討する。

## 2 RBF ネットワークとアンサンブル学習

### 2.1 RBF ネットワーク

生体の神経組織を模した情報処理網であるニューラルネットワークの一種に動径基底関数 (RBF) ネットワークがある。RBF ネットワークは図 1 に示すとおり、入力層、中間層、出力層の三層から構成され、基底関数の線形和によって関数を近似する数理モデルである。ニューロン数と結合重みを適切に定めることで任意の非線形な入出力関係を近似することが可能な universal approximator であることが示されている。

本研究では基底関数にガウス関数を用いる。 $\phi_j(\mathbf{x})$  を  $j$  番目の基底関数とすると、RBF ネットワークの出力は

$$\hat{y} = \sum_{j=0}^m w_j \phi_j(\mathbf{x}), \quad \phi_j(\mathbf{x}) = \exp\left(-\frac{\sum_{i=1}^d (x_i - c_{ij})^2}{2\sigma_j^2}\right) \quad (1)$$

と表わされる。ここで、 $\hat{y}$  は RBF ネットワークの出力、 $m$  は中間層ニューロンの数 (基底関数の数)、 $w_j$  は  $j$  番目のニューロンから出力層への結合重み、 $x_i$  は入力の  $i$  次元の成分、 $c_{ij}$  は  $j$  番目の基底関数の  $i$  次元の成分である。また、 $\phi_0 = 1$  である。

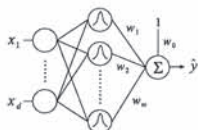


図 1 RBF ネットワーク

一般に学習モデルは、モデル構造を複雑にすればデータへの当てはめが良くなるが、ノイズを含んだデータへの過剰適合 (過学習) が起こり、汎化能力が悪化する。逆にモデルを簡単にすれば過学習を防ぐことができるが、モデルの自由度が損なわれるため対象となる入出力関係を精度よく表現できなくなる [1]。RBF ネットワークにおいては、層数・ニューロン数やニューロンの結合のしかたをどのように決定するかが問題となる。

これまでに、正則化の枠組みや、情報量規準を用いることなどからモデルの複雑さに関するトレードオフを考慮することが検討されている。また、RBF ネットワークの構築を、モデルの精度とパラメータ数あるいは正則化項についての多目的最適化とみなすことで上記のジレンマを解決しようとするアプ

ローチもある [3]。一方で、RBF ネットワークの各種パラメータやネットワーク構造そのものを進化的計算によって学習する手法も提案されている [4]。

### 2.2 アンサンブル学習

近年、機械学習やニューラルネットワークの分野において、アンサンブル学習が注目を集めている。アンサンブル学習は、複数の学習機械の出力を組み合わせることで高い汎化能力を持たせることや複雑な問題を解くことを目指す学習法である。優れたアンサンブルを構築するためには、個々のメンバーがなるべく正確であると同時に、メンバーの多様さが必要であることがわかっている [5]。これを踏まえて、どのようなモデル集合を考えるか、そしてそれらの出力をどのような方法で組み合わせるかがアンサンブル学習における課題となる。

## 3 進化的多目的最適化

多目的最適化問題は、複数の目的関数  $f_m(\mathbf{x}) (m = 1, \dots, M)$  をある制約条件下で最小化 (最大化) する  $d$  次元の決定変数  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$  ( $T$  は転置) を求める問題である。多目的最適化問題においては、目的関数間のトレードオフを考慮するため、「優越」の概念を導入する。

「優越」は 2 つの解の間の関係を表わす。ある解  $\mathbf{x}_1, \mathbf{x}_2$  に対し、

$$\forall i = 1, 2, \dots, M \quad f_i(\mathbf{x}_1) \leq f_i(\mathbf{x}_2) \quad (2)$$

かつ  $\exists j = 1, 2, \dots, M \quad f_j(\mathbf{x}_1) < f_j(\mathbf{x}_2)$

となるとき、 $\mathbf{x}_1$  は  $\mathbf{x}_2$  に優越するという。

そして、どの解にも優越されない解を「パレート最適解」または「非劣解」と呼ぶ。一般にパレート最適解は集合として得られる。多目的最適化問題では、真のパレート最適解より近く、かつなるべく一様に分布した解集合を得ることが要求される。

進化的計算手法により多目的最適化問題を解くところみとして、進化的多目的最適化 (Evolutionary Multi-objective Optimization) が盛んに研究されている。多目的最適化問題を進化的計算によって解くことの最大のメリットは、探索の並列性によって非劣個体を一度に求めることができることにある。

## 4 進化的多目的最適化法による RBF ネットワークのアンサンブル学習

### 4.1 RBF ネットワークの進化的多目的最適化

本研究では 2.1 節で述べたモデルの構造決定に対するひとつのアプローチとして、RBF ネットワークの学習を多目的最適化問題として捉え、進化的多目的最適化法によってさまざまな構造のモデルを求めることを考える。

本研究では、Deb らによる NSGA-II を用いて中間層ニューロン数  $m$  と基底関数の中心  $c_j (j =$



$1, \dots, m$ ) を推定する。NSGA-II は個体の密集度を考慮することによってパレートフロント上の多様性維持を実現しようとするものであり、その有用性は高い [6]。

NSGA-II の染色体は、長さが学習データ数と等しいバイナリ配列とする。各遺伝子は各学習データの位置に対応し、遺伝子の値が 1 ならば対応する位置に基底関数の中心を置き、0 ならば置かないものとする。このようにコーディングすることで、中間層ニューロン数と基底関数の中心位置が表現できる。遺伝的操作としては、トーナメントサイズ 2 のトーナメント選択、一様交叉、ビット反転突然変異を用いる。

上記のコーディングで表現される RBF ネットワークに対し、結合重みを正則化最小二乗法で推定する。すなわち、

$$\mathbf{w} = (\Phi^T \Phi + \alpha \mathbf{E})^{-1} \Phi^T \mathbf{y} \quad (3)$$

によって結合重みが推定される。ここで、 $\mathbf{w} = (w_0, w_1, \dots, w_m)^T$ 、 $\Phi = (\phi_0, \phi_1, \dots, \phi_m)$ 、 $\phi_i = (\phi_i(\mathbf{x}_1), \phi_i(\mathbf{x}_2), \dots, \phi_i(\mathbf{x}_n))^T (i = 0, \dots, m)$ 、 $\mathbf{y} = (y_1, y_2, \dots, y_n)$  である。 $\mathbf{E}$  は単位行列を表す。 $\alpha$  は正則化パラメータである。

基底関数の幅  $\sigma_j (j = 1, \dots, m)$  は、簡単のため一定の値に固定する。

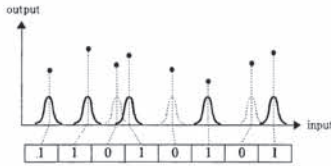


図 2 RBF ネットワークの遺伝子表現

NSGA-II における個体の評価量としては、次の 2 つの値を考える。

1. 中間層ニューロンの数。これによってモデル構造の複雑さを評価する。
2.  $\log MSE$  を小数点第三位で四捨五入したもの。 $MSE$  はデータとモデル出力との平均二乗誤差である。これによってモデルの精度を評価する。

#### 4.2 RBF ネットワークアンサンブル

本研究では、前節の手法により得られた RBF ネットワーク集合からアンサンブルを構築し、パターン分類問題に適用することを考える。

アンサンブルに使用するメンバーについては以下の 4 通りを考える。

1. 最終世代の全個体
2. 最終世代の全非劣個体

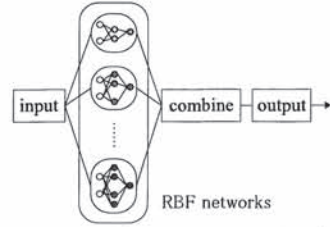


図 3 RBF ネットワークアンサンブル

3. 最終世代の全個体のうち、トレーニングデータに対する正答率が集合の平均値以上の個体
4. 最終世代の全個体から GA によって選択された個体

以下では、それぞれメンバー 1、メンバー 2、メンバー 3、メンバー 4 とよぶ。メンバー 4 を選択する GA は、得られた RBF ネットワークそれぞれをバイナリの遺伝子に対応づけ、値が 1 ならメンバーとして選択、0 なら不選択とするような遺伝子表現をもつ GA とする。個体の評価には、次の disagreement measure の値を RBF ネットワークのすべてのペアについて計算し、それを平均したものを用いる。

$$Dis_{i,j} = \frac{N^{10} + N^{01}}{N^{11} + N^{10} + N^{01} + N^{00}} \quad (4)$$

$Dis_{i,j}$  はモデル  $i, j$  についての disagreement measure である。 $N^{ab} ((a, b) \in \{0, 1\})$  はモデル  $i, j$  がそれぞれ正解 (1) または不正解 (0) となるような解答を出力するデータの数である。これは 2 つのモデルが異なるクラスラベルを出力するデータの割合を示しており、メンバーの多様性の指標としてしばしば用いられる [7]。遺伝的操作としては、トーナメントサイズ 2 のトーナメント選択、一点交叉、ビット反転突然変異を用いる。

出力の組み合わせ方としては、パターン分類に対するアンサンブル学習においてよく用いられる、次の 3 つを考える。

- Simple averaging : 全メンバー出力値の算術平均をアンサンブルの出力とする。
- Majority voting : 最も多くのメンバーが出力したクラスラベルをアンサンブルの出力クラスラベルとする。
- Winner-takes-all : 最も大きな値を出力したメンバーの出力をアンサンブルの出力とする。

以上のメンバー選択法と出力の組み合わせ法の違いによって RBF ネットワークアンサンブルにどのような性能の差異が表れるかを検証するため、数値シミュレーションを行った。パターン分類に関する多くのデータセットが提供されている UCI Machine Learning Repository[8] から、Australian

credit card と Diabetes のデータセットを用いて、10-fold cross validation 法による評価を行った。これらのベンチマーク問題はどちらも 2 クラス分類問題であり、Australian credit card データセットはデータ数が 690 で 14 の特徴量、Diabetes データセットはデータ数が 768 で 8 の特徴量をもつ。

このシミュレーションにおける GA の各パラメータを表 1 に示す。最小二乗法の正則化パラメータ  $\alpha$  は 0.001, 基底関数の幅  $\sigma^2$  は 2.0 としてシミュレーションを行った。提案手法のそれぞれと、ニューラルネットワークのアンサンブル法である Liu らの EENCL[9], Abbass の MPANN[10] に対し、メンバーの個数を表 2 に、テストデータに対する正答率を表 3, 4 にそれぞれ示す。

表 1 GA のパラメータ

	NSGA-II	メンバー選択 GA
世代数	100	100
個体数	30	30
交叉確率	0.8	0.7
突然変異確率	0.05	0.3

表 2 メンバー数 (mem.n はメンバー n を表わす。メンバー 1 は全個体が選ばれるため、メンバー数はすべて 30。EENCL, MPANN はともにメンバー数 25)

	Australian Credit Card			Diabetes		
	mem.2	mem.3	mem.4	mem.2	mem.3	mem.4
Mean	19.5	18.8	8.1	8.9	19.2	8.2
S. D.	2.2	1.1	2.2	1.2	2.4	2.6
Max	23.0	21.0	13.0	11.0	22.0	13.0
Min	17.0	17.0	5.0	7.0	14.0	4.0

表 3 Australian Credit Card の結果。SA, MV, WTA はそれぞれ Simple Averaging, Majority Voting, Winner-Takes-All を表わす。

		mem.1	mem.2	mem.3	mem.4	EENCL	MPANN
		SA	Mean	0.861	0.864	0.870	0.868
	S. D.	0.029	0.030	0.032	0.032	0.043	0.039
	Max	0.913	0.913	0.913	0.913	0.928	0.913
	Min	0.797	0.797	0.797	0.797	0.797	0.797
MV	Mean	0.862	0.868	0.871	0.868	0.862	0.857
	S. D.	0.034	0.030	0.033	0.031	0.049	0.039
	Max	0.913	0.913	0.913	0.913	0.928	0.913
	Min	0.783	0.797	0.797	0.797	0.783	0.812
WTA	Mean	0.858	0.865	0.861	0.868	0.858	0.865
	S. D.	0.035	0.027	0.031	0.030	0.044	0.028
	Max	0.899	0.899	0.899	0.899	0.913	0.913
	Min	0.812	0.812	0.783	0.797	0.797	0.812

表 4 Diabetes の結果

		mem.1	mem.2	mem.3	mem.4	EENCL	MPANN
		SA	Mean	0.764	0.764	0.768	0.774
	S. D.	0.045	0.051	0.048	0.048	0.039	0.032
	Max	0.855	0.855	0.855	0.855	0.828	0.840
	Min	0.711	0.684	0.697	0.697	0.703	0.723
MV	Mean	0.772	0.776	0.772	0.778	0.764	0.779
	S. D.	0.048	0.053	0.052	0.047	0.042	0.033
	Max	0.868	0.882	0.868	0.868	0.828	0.840
	Min	0.711	0.711	0.697	0.711	0.688	0.723
WTA	Mean	0.750	0.755	0.775	0.776	0.779	0.777
	S. D.	0.050	0.053	0.047	0.046	0.045	0.037
	Max	0.829	0.842	0.855	0.855	0.844	0.840
	Min	0.658	0.645	0.697	0.697	0.703	0.723

どちらのデータセットにおいても、最終世代での全個体を用いるよりも、部分集合をうまく選択することで正答率を上げることができ、EENCL, MPANN と同等以上の性能を示した。なかでも、メ

ンバーの多様性の観点から GA で選択したメンバー 4 は、小さなメンバー数で比較的安定した結果を出している。

## 5 おわりに

本研究では、進化的多目的最適化法を用いて得られた RBF ネットワークの集合からアンサンブルを構築する方法について検討した。本手法では、構造の複雑さとモデルの精度の 2 つの評価基準にもとづいた NSGA-II の染色体に RBF ネットワークの構造がコーディングされ、これによって学習された RBF ネットワークの集合からアンサンブルが構成される。数値シミュレーションでは、アンサンブルメンバーの選択手法やアンサンブル手法を適切に定めることで従来手法と同等以上の結果を得ることができた。今後は、より積極的に多様性を生み出すメカニズムについても検討したい。

## 参考文献

- [1] Oliver Nelles, *Nonlinear System Identification*, Springer (2001)
- [2] Amanda J. C. Sharkey, On Combining Artificial Neural Nets, *Connection Science*, Vol.8, pp.299-313 (1996)
- [3] Yaochu Jin, Tatsuya Okabe and Bernhard Sendhoff, Neural Network Regularization and Ensembling Using Multi-objective Evolutionary Algorithms, *Proceedings of Congress on Evolutionary Computation*, pp.1-8 (2004)
- [4] Yunfei Bai and L. Zhang, Genetic Algorithm Based Self-Growing Training for RBF Neural Networks, *Proceedings of International Joint Conference on Neural Networks 2002*, pp. 840-845 (2002)
- [5] Thomas G. Dietterich, Ensemble Methods in Machine Learning, In *Proceedings of the First International Workshop on Multiple Classifier Systems 2000*, Springer Verlag, pp.1-15 (2000)
- [6] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal and T.Meyarivan, A Fast and Elitist Multiobjective Genetic Algorithm : NSGA-II, *IEEE Transactions on Evolutionary Computation*, Vol.6, No.2, pp.182-197 (2002)
- [7] Ludmila I. Kuncheva and Christopher J. Whitaker, Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy, *Machine Learning*, Vol.51, pp.181-207 (2003)
- [8] <http://www.ics.uci.edu/~mllearn/>
- [9] Yong Liu, Xin Yao and Tetsuya Higuchi, Evolutionary Ensembles with Negative Correlation Learning, *IEEE Transactions on Evolutionary Computation*, Vol.4, No.4, pp.380-387 (2000)
- [10] Hussein. A. Abbass, Pareto Neuro-Evolution : Constructing Ensemble of Neural Networks Using Multi-objective Optimization, *The IEEE Congress on Evolutionary Computation 2003*, Vol.3, pp.2074-2080 (2003)