

スイッチング機構を持つ一般化学習ネットワークの汎化能力

韓, 敏

九州大学大学院システム情報工学研究科電気電子システム工学専攻 : 博士後期課程

平澤, 宏太郎

九州大学大学院システム情報工学研究科電気電子システム工学専攻

大林, 正直

九州大学工学部電気情報工学教室

藤田, 寛文

九州大学総合理工学研究科エネルギー変換工学専攻 : 修士課程

<https://doi.org/10.15017/1524285>

出版情報 : 九州大学大学院システム情報科学紀要. 2 (2), pp.289-295, 1997-09-26. Graduate School of Information Science and Electrical Engineering, Kyushu University

バージョン :

権利関係 :

スイッチング機構を持つ一般化学習ネットワークの汎化能力

韓 敏* · 平澤宏太郎** · 大林正直*** · 藤田寛文†

Generalization Ability of Universal Learning Network with Switching Mechanism

Min HAN, Kotaro HIRASAWA, Masanao OHBAYASHI and Hirofumi FUJITA

(Received June 23, 1997)

Abstract: Universal Learning Network(ULN), which can model and control large scale complicated systems naturally, consists of nonlinearly operated nodes and multi-branches that may have arbitrary time delays including zero or minus ones. It can therefore be applied to many kinds of systems which are difficult to be expressed by ordinary first order difference equations with one sampling time delay. In order to improve the generalization ability of ULN, in this paper an approach to optimize the structure of networks are presented, where not only parameters but also time delays in the ULN are adjusted. In the proposed method, both the compactness and error criterion function of the ULN are improved. Simulation results of a nonlinear system identification show that better performance can be obtained by the proposed method than the conventional method using only the parameter optimization.

Keywords: Generalization ability, Search, Time delay, Training, Universal Learning Network

1. はじめに

ニューラルネットワークに関する研究は80年代後半から第二次ブームになり、コンピュータの進歩を背景に、制御工学や情報工学への応用が盛んになされるようになった。以来、システム同定や制御の方でも数多くの研究がなされ、その研究成果が数多く報告されている。

ニューラルネットワークを応用するにあたり注意すべき点が幾つかある。一つは、ネットワークの表現能力の問題であり、一つは、汎化能力の問題である。ネットワークの表現能力に関しては、著者等は、その能力を高めるべく、ノード間を複数のブランチで接続可能とし、且つ各ブランチに任意の遅れ時間を設定可能とした一般化学習ネットワークを既に提案している¹⁾²⁾。しかしながら、表現能力が高まるにつれ、汎化能力をいかに維持するかが重要になってくる、即ち、表現能力が高まると、その汎化能力を維持するには、学習データ量を多くする必要がある。しかしながら、学習データが数多く手に入らない場合は、ネットワークの表現能力を減小させる、即ち、その自由度を少なくする必要がある。

このニューラルネットワークの汎化能力を高める方策として、種々の方法が提案されている。例えば、情報量基準に基づく方法、ネットワークを簡素化することによる方

法等がある。これらについては、文献3)に詳しくまとめられている。

著者等は、汎化能力を高める方策として、ネットワークブランチにスイッチング機構を設け、ネットワークのコンパクト性とモデリング誤差最小化を同時に考慮した同定法をすでに提案している⁵⁾。

しかしながら、文献5)では、ネットワーク内の遅れ時間は固定した状態で、ネットワークのコンパクト化がなされており、遅れ時間の最適化は考慮されていない。本論文では、文献5)の方法に更に、最適遅れ時間の探索を考慮したネットワークのコンパクト化とモデリング誤差の最小化を同時に考慮した同定法を提案する。シミュレーションにより、文献5)と比較し遅れ時間の探索を考慮した本論文提案の方が、同定誤差も少なく、かつ汎化能力も優れていることを示す。

2. 一般化学習ネットワーク

一般化学習ネットワーク¹⁾²⁾とは、物理現象、社会現象、生命現象等の非線形な複雑大規模システムを自然な形でモデル化し制御するための枠組である。ネットワークによる自然な形でのモデル化とは、連続微分可能な任意の非線形処理ノードが、任意のサンプル遅れ時間を介して任意にマルチ接続された離散時間形のネットワークでシステムをモデル化すると、遅れ時間の設定又は把握が適切に出来ることをいう。特にブランチ間の遅れ時間を負の値に設定すると予測機構を導入したシステムのモデル化も可能になる。

平成9年6月23日受付

* 電気電子システム工学専攻博士後期課程

** 電気電子システム工学専攻

*** 工学部電気情報工学教室

† 総合理工学研究科エネルギー変換工学専攻修士課程

2.1 スイッチング機構を持つ一般化学習ネットワーク (ULN) の構成

ULN はFig.1 に示すように、連続微分可能な任意の非線形処理ノード N_j が、任意のサンプル遅れ時間 $D_{ij}(p)$ を介して任意にマルチ接続された離散時間型のネットワークである。

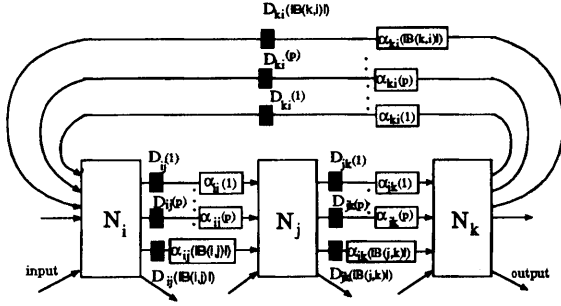


Fig.1 Structure of ULN with multi-branches and switching mechanisms

一般化学習ネットワークの基本式は

$$h_j(t) = O_j(\{h_i(t - D_{ij}(p)) | i \in JF(j), p \in B(i, j)\}, \{\lambda_m(t) | m \in M(j)\}), j \in J, t \in T \quad (1)$$

となる。
ここで、

- $h_j(t)$: j ノードの t 時刻の出力値,
- $\lambda_m(t)$: m パラメータの t 時刻の値,
- $r_n(t)$: n 外部入力変数の t 時刻の値,
- O_j : j ノードの非線形関数,
- $D_{ij}(p)$: i ノードから j ノードへの p 番目のブランチの遅れ時間,
- $JF(j)$: j ノードへ接続するノードの番号の集合,
- $JB(j)$: j ノードから接続するノードの番号の集合,
- $B(i, j)$: i ノードから j ノードへ接続するブランチ番号の集合,
- $N(j)$: j ノードへ接続する外部入力変数番号の集合,
- N : 外部入力変数番号の集合,
- $M(j)$: j ノードへ接続するパラメータ番号の集合,
- M : パラメータ番号の集合,
- J : ノード番号の集合,
- T : 時刻の集合.

である。特に N_j ノードの非線形関数が $[0,1]$ のシグモイド関数 $f(\alpha_j(t))$ の場合(1)式は

$$h_j(t) = f(\alpha_j(t)) = \frac{1}{1 + e^{-\psi_j \alpha_j(t)}} \quad (2)$$

$$\alpha_j(t) = \sum_{i \in JF(j)} \sum_{p \in B(i, j)} h_{ij}(p) \alpha_{ij}(p) h_i(t - D_{ij}(p)) + \theta_j \quad (3)$$

である。

ここで、スイッチング機構は下記のシグモイド関数で表現し、

$$\alpha_{ij}(p) = \frac{1}{1 + e^{-\psi \beta_{ij}(p)}} \quad (4)$$

$\beta_{ij}(p)$ を学習パラメータとして $h_{ij}(p)$ と同時に学習を行う。 ψ をはじめ小さくしておき、学習が進むにつれて、 ψ を十分大きくすることにより、 $\beta_{ij}(p)$ が 0 より大きければ $\alpha_{ij}(p)$ の値は 1 となり、ブランチを接続、 $\beta_{ij}(p)$ が 0 より小さければ $\alpha_{ij}(p)$ の値は 0 となり、ブランチを削除というスイッチング機構を実現できる。

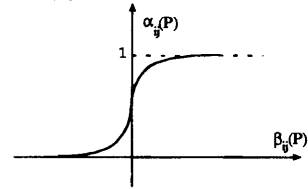


Fig.2 Switching function

又、コンパクトなネットワークを形成するために、総和評価指標 L を次のようにする。

$$L = E + R_\alpha \sum_i \sum_j \sum_p (\alpha_{ij}(p))^2 \quad (5)$$

$$E = E(\{h_r(s)\}, \{\lambda_m(s)\})$$

ここで、

- E : システムの評価指標,
- R_α : スイッチング機構の重み係数,
- $\alpha_{ij}(p)$: i ノードから j ノードへの p 番目のブランチのスイッチング機構.

2.2 スイッチング機構を持つ一般化学習ネットワークの学習アルゴリズム

(1)式で定式化された ULN 上で(5)式の評価指標を局所最小とするパラメータを学習により求めるアルゴリズムを考える。パラメータの学習アルゴリズムは文献 1), 2) によると、

$$\lambda_m(t) \leftarrow \lambda_m(t) - \gamma \frac{\partial L}{\partial \lambda_m} \quad (6)$$

$$\beta_{ij}(p) \leftarrow \beta_{ij}(p) - \gamma \frac{\partial^\dagger L}{\partial \beta_{ij}(p)} \quad (7)$$

$$\frac{\partial^\dagger L}{\partial \lambda_m} = \sum_{t' \in T} \sum_{d \in JD(\lambda_m)} \left[\frac{\partial h_d(t')}{\partial \lambda_m} \delta(d, t') \right] + \frac{\partial L}{\partial \lambda_m} \quad (8)$$

$$\frac{\partial^\dagger L}{\partial \beta_{ij}(p)} = \sum_{t' \in T} \left[\frac{\partial h_j(t')}{\partial \beta_{ij}(p)} \delta(j, t') \right] + \frac{\partial L}{\partial \beta_{ij}(p)} \quad (9)$$

$$\delta(j, t) = \sum_{k \in JB(j)} \sum_{p \in B(j, k)} \left[\frac{\partial h_k(t + D_{jk}(p))}{\partial h_j(t)} \times \delta(k, t + D_{jk}(p)) \right] + \frac{\partial L}{\partial h_j(t)} \quad (10)$$

$$j \in J, t \in T$$

となる。

また, $\frac{\partial^\dagger L}{\partial \lambda_m}$, $\frac{\partial^\dagger L}{\partial \beta_{ij}(p)}$ は Werbos⁴⁾ の提案したオーダード微分であり, 他の変数を固定して, λ_m , $\beta_{ij}(p)$ が変動した場合の評価指標 E の変動分を表している。

ただし,

$JD(\lambda_m)$: 出力が λ_m で偏微分可能なノード番号の集合。

処理ノードの非線形関数が $[-A, A]$ のシグモイド関数の場合, 式(8),(9),(10)は次のようになる。

$$h_j(t) = A \frac{1 - e^{-\alpha_j}}{1 + e^{-\alpha_j}} \quad (11)$$

$$\alpha_j = \sum_{i \in JF(j)} \sum_{p \in B(i, j)} h_{ij}(p) \alpha_{ij}(p) \times h_i(t - D_{ij}(p)) + \theta_j \quad (12)$$

$$\frac{\partial h_j(t)}{\partial h_{ij}(p)} = \frac{A}{2} \left\{ 1 - \left(\frac{h_j(t)}{A} \right)^2 \right\} \alpha_{ij}(p) \times h_i(t - D_{ij}(p)) \quad (13)$$

$$\frac{\partial h_j(t)}{\partial \beta_{ij}(p)} = \frac{A}{2} \left\{ 1 - \left(\frac{h_j(t)}{A} \right)^2 \right\} h_{ij}(p) \times h_i(t - D_{ij}(p)) \psi \times \alpha_{ij}(p) (1 - \alpha_{ij}(p)) \quad (14)$$

$$\frac{\partial L}{\partial \beta_{ij}(p)} = 2R_\alpha \psi \alpha_{ij}(p)^2 (1 - \alpha_{ij}(p)) \quad (15)$$

$$\frac{\partial h_k(t + D_{jk}(p))}{\partial h_j(t)} = \frac{A}{2} \left\{ 1 - \left(\frac{h_k(t + D_{jk}(p))}{A} \right)^2 \right\} \times h_{jk}(p) \alpha_{jk}(p) \quad (16)$$

3. ニューラルネットワークの汎化能力

ニューラルネットワークの汎化能力とは, 学習で得られたネットワークが, 未学習の入力データに対して妥当な出力を生成できるかどうかという意味である。汎化能力はネットワークの性質の中でもとくに重要な問題に位置づけられる。

ネットワークの自由度(ノード数やブランチ数)が十分でない場合には, ネットワークの表現能力が十分でなく, 自由度を上げていくことにより, 評価指標の値は改善される方向に向かうことになる。同時に未学習データに対しても評価指標の値はよくなることが考えられる。しかしながら, ネットワークの自由度が大きくなりすぎると, 学習用データに対しては評価指標の値はより減少させることができるが, 未学習データに対しては必ずしもよい結果を示さなくなる。また, 同じ自由度であっても学習の繰り返し回数が多過ぎると同様の現象が起こる。いわゆる過学習という問題もある。

汎化能力を向上させる方策として, 基本的に学習データを豊富にするか, ネットワークの自由度を減少させる(ネットワークサイズのコンパクト化)の二つの方法がある。本論文では, 後者の方法により, 汎化能力を確保するとともに同定誤差も最小とするネットワーク構成を検討する。この時, 次節に述べる方法にてネットワーク内の遅れ時間の最適化もあわせて実現する。

4. 遅れ時間のランダム探索

ここでは, 一般化学習ネットワークのノード間の遅れ時間の探索の方法を示す。

一般化学習ネットワークの特徴であるノード間の遅れ時間が任意に設定可能であるというを用い, システムを同定するとき, システムの遅れ時間が既知の場合はシステムに対応した遅れを設定した方が, 1 サンプル遅れの従来のニューラルネットワークモデルよりも優れていることがいえる。しかし, システムの遅れ時間は多くの場合は未知であるので, あらかじめ遅れ時間を設定することはできない。そこで, 重みなどのパラメータと同時に遅れ時間を探索することにより, システムに対応した遅れ時間を持つネットワークが実現できるのではないかと考えられる。

最適遅れ時間探索のアルゴリズムを次に示す。

Step 1. 各ブランチの遅れ時間を乱数を用いて $[1, 10]$ の範囲の整数にて初期値設定する。

Step 2. ネットワークのパラメータをN回学習する.

Step 3. 探索するブランチをランダムに一つ選ぶ. 遅れ時間を全範囲で探索し, 誤差が最小になる遅れ時間を選択する. 誤差の改善があれば探索を終了, なければ他のブランチ選択して同様のことを行い, 誤差の改善がなされるか, 最大許容回数になるまで繰り返す.

Step 4. Step 2., Step 3. をQ回繰り返す.

すなわちパラメータの学習は $N \times Q$ 回, 遅れ時間の探索はQ回行うことになる.

5. 非線形システムの同定とシミュレーション

不連続関数で表現される非線形要素は, 実際の物理的システムではよくみかけられる. 本文には, 非線形要素としてリレー特性を選び, 非線形要素 g_N が線形要素 g_L と直列に接続された非線形システムの同定問題について考える (Fig.3参照).

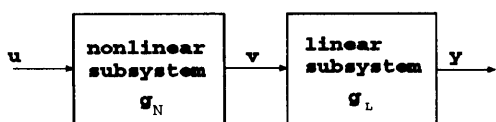


Fig.3 Cascade connection of nonlinear and linear subsystem

システムの入力 $u(k)$ と出力 $y(k)$ を(17)式に示す.

$$y(k+1) = \begin{cases} 1.34y(k) - 0.277y(k-2) \\ -0.80y(k-4) + 0.01, & u(k) \geq 0 \\ 1.34y(k) - 0.277y(k-2) \\ -0.80y(k-4) - 0.01, & u(k) < 0 \end{cases} \quad (17)$$

又, Fig.4 に同定すべきシステムの入力波形と出力波形を示す.

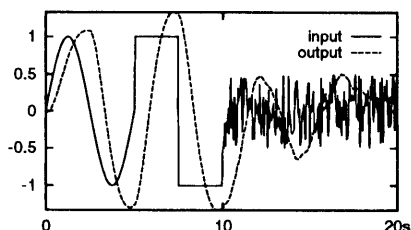


Fig.4 Input and output data for training(case 0)

Fig.5 ~ Fig.6は汎化能力を検証するための入力波形と出力波形である. これらはFig.4の三つの出力パターンをそれぞれ順序を変えて形成した.

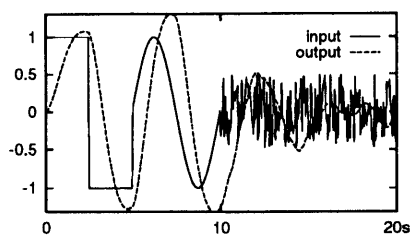


Fig.5 Input and output data for testing the generalization ability(case 1)

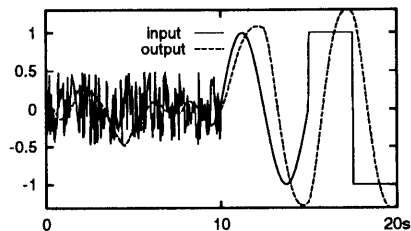


Fig.6 Input and output data for testing the generalization ability(case 2)

入力信号は以下のように設定した.

$$u(k) = \begin{cases} \sin(\frac{\pi}{50}k), & 0 \leq k < 100 \\ 1.0, & 100 \leq k < 150 \\ -1.0, & 150 \leq k < 200 \\ \text{uniform random,} & 200 \leq k < 400 \\ \text{number in}(-0.5, 0.5), & \end{cases}$$

シミュレーションは, 遅れ時間の探索を行わない場合と行う場合について実施した. 遅れ時間の探索を行わない場合は, 遅れ時間を全て1に設定した場合(case a)と $[1, \dots, 10]$ の範囲でランダムに設定した場合(case b)の2ケース, 行う場合は, その探索回数を50回(case c), 500回(case d), 5000回(case e) の3ケースの計5ケースについて行った.

シミュレーションの条件をTable 1に示す. 学習パラメータ λ_m の初期値は, 乱数を変えて3ケース, 遅れ時間の初期値も乱数を変えて3ケース, R_α は0.0, 0.1, 0.5の3ケースについて行った.

その結果をTable 2 ~ Table 4に示す.

Table 2 ~ Table 4に R_α の値ごとの学習終了時の接続ブランチ数及び同定誤差の平均値を示す. これらより, R_α が大きい程接続ブランチ数は少ないが, 当然同定誤差は大きくなっていることがわかる. R_α が同じ値については, ウェイト学習のみの場合より遅れ時間の探索を行った時の方が同定誤差が少ない. また, 探索回数は多い方がシステムにあった最適遅れ時間の探索が成功し, 同定誤差が少なくなっていることがわかる.

Table 1. Simulation conditions

number of nodes	J=5
number of branches between nodes	1
nonlinear function	$f(x) = A \frac{1-e^{-x}}{1+e^{-x}}$ A=1.5
initial value of parameter	random numbers in (-1.0,1.0)
λ_m	random numbers in (-1.0,1.0)
$\beta_{ij}(p)$	0.3 (fully connected)
learning coefficient of $\lambda_m, \beta_{ij}(p)$	$\gamma_\lambda=0.00002, \gamma_\beta=0.0002$
identification error E coefficient R_α	root square error 0.0, 0.1, 0.5
number of learning of parameters	500000
number of time delay search	50,500, 5000
ϕ in switching function	increase from 20 to 5000

Table 2. Identification results (weight coefficient $R_\alpha = 0.0$)

case	a	b	c	d	e
parameter training number after search of delay	500000	500000	10000	1000	100
search number	0	0	50	500	5000
residual branches	25	25	25	25	25
average error $\times (10)^{-3}$	4.42	2.78	2.97	3.21	3.09

Table 3. Identification results (weight coefficient $R_\alpha = 0.1$)

case	a	b	c	d	e
parameter training number after search of delay	500000	500000	10000	1000	100
search number	0	0	50	500	5000
residual branches	13	8	9	8	9
average error $\times (10)^{-3}$	6.33	5.25	5.18	3.68	3.78

Table 4. Identification results (weight coefficient $R_\alpha = 0.5$)

case	a	b	c	d	e
parameter training number after search of delay	500000	500000	10000	1000	100
search number	0	0	50	500	5000
residual branches	7	7	7	7	5
average error $\times (10)^{-3}$	10.4	9.0	7.85	7.65	5.5

Fig.7, Fig.8, Fig.9は, R_α が異なる場合について, 横軸を, Fig.4の学習データである case0, Fig.5, Fig.6の未学習データである case1, case2とし, 縦軸は, それぞれの場合の同定誤差の平均値である. ここで, 平均値の意味はパラメータの初期値(3) \times 遅れ時間の初期値(3)の各シミュレーションの誤差の最大値及び最小値をカットし7つのケースの平均値をいう. case1, case2の未学習データ, 即ち, 汎化能力についてみてみるとコンパクト化を行った, 即ち

R_α が零でない場合で(d)(e)と探索回数が多い程誤差が少ない, 即ち, 汎化能力に優れていることがわかる.

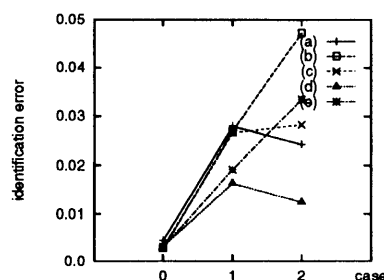


Fig.7 Average identification error ($R_\alpha = 0.0$)

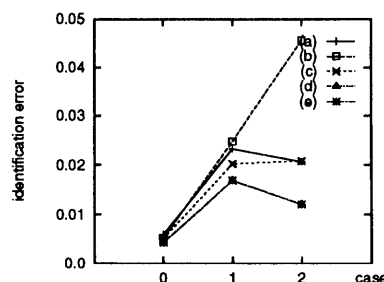


Fig.8 Average identification error ($R_\alpha = 0.1$)

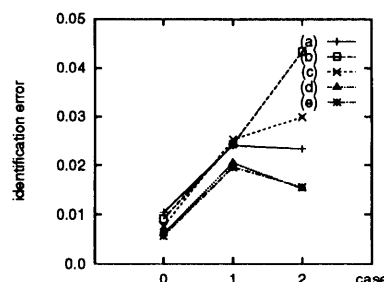


Fig.9 Average identification error ($R_\alpha = 0.5$)

Fig.10 と Fig.11 は $R_\alpha = 0.0$ の場合, Fig.12 と Fig.13 は $R_\alpha = 0.1$ の場合, Fig.14と Fig.15 は $R_\alpha = 0.5$ の場合で, それぞれ前者が遅れ時間を探索なし (Table 2 ~ Table 4 における case b), 後者が遅れ時間探索回数が5000回 (Table 2 ~ Table 4 における case e)の時のシミュレーション結果である. これらで探索なしと探索有りの場合を比較すると学習データ (case 0), 未学習データ (case 1, case 2)とも, 又各 R_α について探索有りの方が同定誤差が少なくなかつ汎化能力も優れていることがわかる. 探索なしの場合の R_α の違い及び探索回数が5000回の場合の R_α の違いについては, 両者とも, 学習データ (case 0) に対しては, コンパクト化をしない場合 ($R_\alpha = 0.0$) の方がコンパクト化した場合 ($R_\alpha = 0.1, 0.5$) より誤差が少ない. これは, 学習データに対しては, 自由度が大きい方が誤差が少なくなっている場合である. 一方, 未学習データに対しては, R_α が大きい方が誤差は少なく汎化能力が優れている.

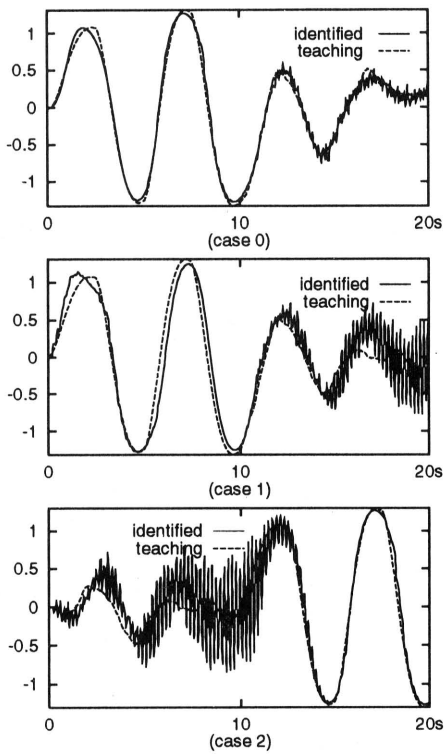


Fig.10 Results of identified signal and teaching signal($R_\alpha = 0.0$. The number of search is 0.)

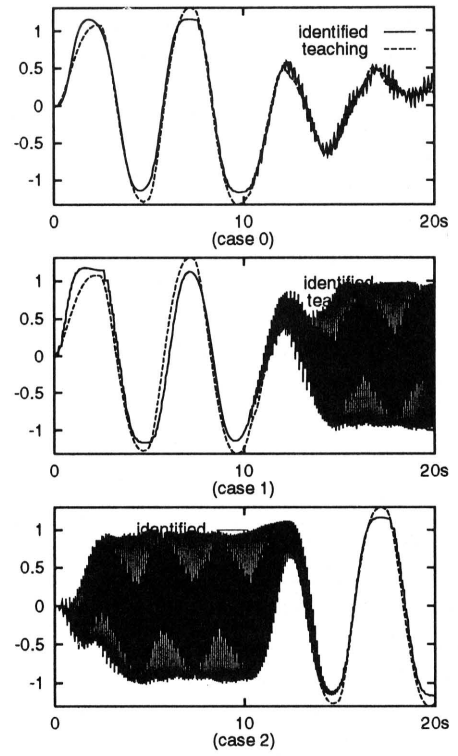


Fig.12 Results of identified signal and teaching signal($R_\alpha = 0.1$. The number of search is 0.)

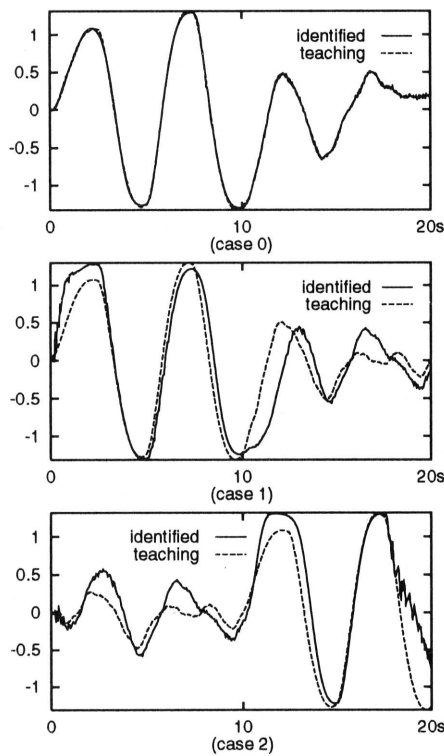


Fig.11 Results of identified signal and teaching signal($R_\alpha = 0.0$. The number of search is 5000.)

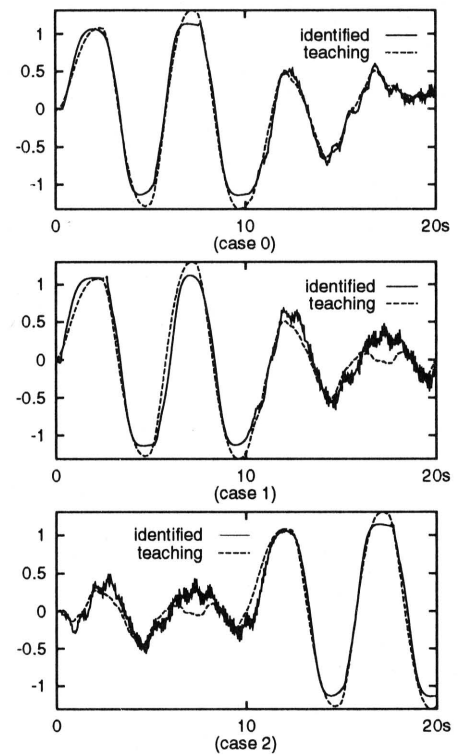


Fig.13 Results of identified signal and teaching signal($R_\alpha = 0.1$. The number of search is 5000.)

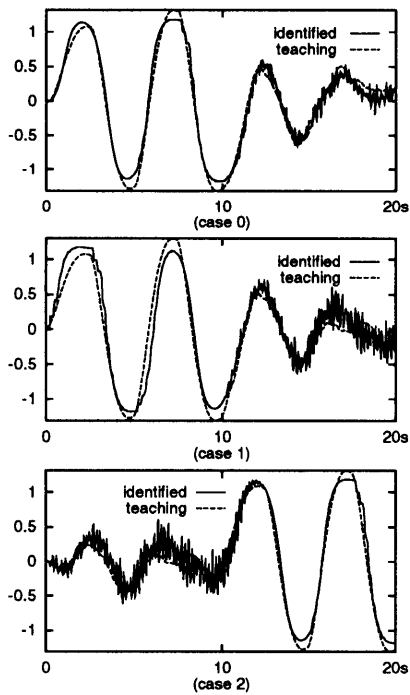


Fig.14 Results of identified signal and teaching signal($R_\alpha = 0.5$. The number of search is 0.)

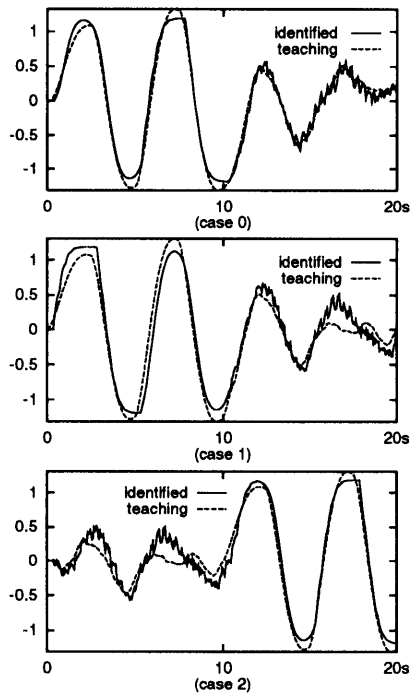


Fig.15 Results of identified signal and teaching signal($R_\alpha = 0.5$. The number of search is 5000.)

以上の結果より、十分な遅れ時間の探索は、遅れ時間の探索のない場合と比較し、表現能力にすぐれていると言える。更に、スイッチング機構のネットワークコンパクト化をも合わせて行うことにより、汎化能力も向上させることができるということが明らかになった。

6. 結 論

本論文では、ネットワークの表現能力及び汎化能力の向上策として、ノード間ブランチに設定する最適遅れ時間の探索とスイッチング機構によるネットワークのコンパクト化と同定誤差の最小化を同時に実現する新たなシステム同定法を提案した。非線形システムの同定と汎化能力についてのシミュレーションを行い、学習データに対しては、表現能力が向上し、未学習データに対しては、優れた汎化能力を有するシステムが提案法により実現可能であることを示した。

参 考 文 献

- 1) Hirasawa K., Ohbayashi M., Murata J. : *Universal Learning Network and Computation of its Higher Order Derivatives*, Proc. of 1995 IEEE International Conference on Neural Networks, pp1273-1277,1995
- 2) Hirasawa K., Ohbayashi M., Koga M. and Harada M. : *Forward Propagation Universal Learning Network*, Proc. of 1996 IEEE International Conference on Neural Networks, pp 353-358,1996
- 3) 喜多 一: ニューラルネットワークの汎化能力, システム/制御/情報/, Vol.36 No.10.pp625-633,1992
- 4) Werbos.P.: *Beyond regression: New Tools for Prediction and Analysis in the Behavior Science*, Ph.D. dissertation, Harvard University,1974
- 5) Han M., Hirasawa K., Ohbayashi M. and Fujita H: *Modeling Nonlinear Dynamic Systems using Learning Network with Filtering Mechanism*, Trans.IEE of Japan, Vol.117-c, No.9(1997)
- 6) Han M., Hirasawa K., Ohbayashi M. and Fujita H: *Modeling Dynamic Systems using Universal Learning Network*, Proc. of IEEE International Conference on Systems, Man and Cybernetics, pp1172-1177,1996
- 7) Han M., Hirasawa K., Ohbayashi M. and Fujita H : *Generalization Ability of Modeling Dynamic Systems Using Universal Learning Network*, 11th IFAC Symposium on system Identification,1997