

階層型ニューラルネットワークへの入力の選定法

村田, 純一

九州大学大学院システム情報科学府電気電子システム工学専攻

中園, 徹

九州大学大学院システム情報科学府電気電子システム工学専攻 : 修士課程

平澤, 宏太郎

九州大学大学院システム情報科学府電気電子システム工学専攻

<https://doi.org/10.15017/1498360>

出版情報 : 九州大学大学院システム情報科学紀要. 3 (2), pp.219-224, 1998-06-22. 九州大学大学院システム情報科学研究所

バージョン :

権利関係 :

階層型ニューラルネットワークへの入力を選定法

村田純一*・中園徹**・平澤宏太郎***

Input Variable Selection for Multi-Layer Neural Networks

Junichi MURATA, Toru NAKAZONO and Kotaro HIRASAWA

(Received June 22, 1998)

Abstract: A method is proposed for selecting relevant input variables to multi-layer neural networks. A minimal set of inputs is selected which is necessary to obtain a network with a good generalization ability and some insight into the input-output relationship. The inputs of network are selected automatically by a combination of constructive and destructive algorithms. The constructive algorithm starts with a minimal input set and adds new inputs if necessary, while the destructive algorithm deletes unnecessary inputs. The main issue addressed here is the measure of input significance used in the constructive algorithm. Some measures are proposed based on mutual information and linear correlation paying much attention to the structural constraint imposed on the networks. The experimental results show that the measures are valid and that the derived network with the selected inputs has a good generalization ability.

Keywords: Neural networks, Input variable selection, Generalization ability, Interpretability

1. はじめに

本論文では、階層型ニューラルネットワークへの適切な入力の選定を行う方法を提案する。

ニューラルネットワーク(NN)は、十分なノード数を備えていれば与えられたデータを正しく表現でき、ある程度の汎化能力を持つ。しかし、汎化能力を向上させるには、NNの構造を適切に定めることが必要であり、筆者らも内部構造決定の方法を提案している¹⁾。

ここでは、NNの働きが鮮明になる入力に着目した一種の構造決定を行う。出力を説明するのに必要な入力を自動的に選定することにより、汎化能力を高めるばかりでなく、入出力関係を人間が解釈しやすいものにもできる。

2. 入力変数選択による一般性の高いNNの構築

NNはそのままでは十分な汎化能力を持っていないし、ブラックボックスと言われるNNの働きを複数のパラメータ値から人間が解釈するのは容易ではない。そこで、NNの働きが鮮明になる入出力関係に着目する。一般に、考慮する入力変数の個数を増せば学習データについての誤差は減少するが、一方で汎化能力は低下する。これは、学習データが、データ数が有限で誤差やノイズを含んでいる、いわば不完全な情報源であるためである。特に、ネットワークの自由度が高い、すなわち未定パラメータ

(重み)の個数が多いと、情報の不完全さゆえにパラメータの値が適切に決められないという不適切な問題となる。また自由度が高すぎると、データの持つ情報が複数のパラメータに分散されて表現されるため、パラメータ値から人間に解釈可能な情報を読み取るのは困難となる。出力を説明するのに必要な入力のみを選ぶことにより、不要な入力による悪影響を避け、一般性の高いNNを得ることができる。また、出力に関わりの大きい入力は何であるかを知ることは、単に数値的な誤差が小さくなるだけでなく、対象とする入出力関係についての有用な情報を得ることになり解釈可能性の向上にも役立つ。

数多くある入力変数から適切なものをネットワークの入力とすることが必要である。入力変数の候補が多い場合、その組合せも膨大となるので全ての組合せを評価するのは実現不可能であり、効率的な探索方法が望まれる。その方法には、遺伝的アルゴリズムのように試行結果に基づいて次の探索点を発見する方法、全ての入力変数を組み入れたネットワークから、不要な入力を取り除いていく削除的方法、入力変数と出力変数との関係の強さを相互情報量²⁾などによって評価し、関係の強い入力変数から順にネットワークに追加していく生成的方法がある。

ここでは、最小の入力ノード数から始めて生成的方法により逐次入力を追加していき、その後入出力感度による削除的方法^{3)~5)}を試す複合的方法を提案する。生成的方法では教師データに対する表現能力を向上させるために必要以上に大きいネットワークになる傾向がある。そこで、生成的方法でネットワークを構成した後、削除的方法によって余分な入力を除く。一方、削除的方法だけ

平成10年6月22日受付

* 電気電子システム工学専攻

** 電気電子システム工学専攻修士課程

*** 電気電子システム工学専攻

を用いた場合も、ヒルクライム的な方法のため入力変数の候補が多い場合、構造上の局所解に陥りやすく必要以上に大きいネットワークになる。生成的と削除的の両方を組み合わせた方法によって入力を選定することで入力ノード数の少ない NN を構築する。

本研究では、生成的方法において適切な入力を選ぶための尺度の提案が主な目的である。ネットワークに実際に入力を加える前に、入力情報の有用性を測ることができる尺度が必要であり、この尺度として、 NN の構造を考慮したものを提案する。

対象とする NN は階層型ネットワークであり、学習則は BP(Back Propagation)法である。

3. 相互情報量を用いた従来の生成的方法

3.1 方法

NN の働きは、空間的・時間的なパターンの非線形変換である。そこで入力の重要性尺度として、入出力関係が非線形な場合でも取り扱い可能な相互情報量を用いることが提案されている^{2)~5)}。

相互情報量とは X と Y の 2 つの事象がある時、もともと X が持っている乱雑さがもう一方の Y を知ってどれだけ減少するかを表す量である。つまり、 Y を知ることで X の取る値をどの程度まで推測できるかを表す量であるといえる。

相互情報量 $MI(X, Y)$ は Y のエントロピー $H(Y)$ と条件付きエントロピー $H(Y|X)$ を用いて次式で定義される。

$$MI(X, Y) = H(Y) - H(Y | X) \tag{1}$$

この相互情報量を用いて、入力情報 u_h の重要性尺度 $I_m(u_h; d_1, \dots, d_r)$ を以下のように定める。

$$I_m(u_h; d_1, \dots, d_r) = \sum_{j=1}^r MI(u_h, d_j) \tag{2}$$

ここで、 h は入力ノード番号、 j は出力ノード番号、 r は出力ノード数、 u_h は入力変数、 d_j は出力ノードに与えられる教師信号の変数である。

この I_m は入力ノードに入る値 u_h が出力ノードに与えられる教師信号の値 d_1, \dots, d_r の説明にどれだけ有効かを表している。

以下に生成的方法の入力選定手順を示す。

1. $I_m(u_h; d_1, \dots, d_r)$ をすべての $u_h (h = 1, \dots, N)$ について計算し、その最大値を与える u_h を u_H とする。
2. 入力として u_H を用いる 1 入力 r 出力の NN を構成し、学習を行う。得られたネットワークの出力を y_1, \dots, y_r とし、この誤差を $\tilde{d}_j = d_j - y_j$ とする。 ($j = 1, \dots, r$)

3. u_h の重要性尺度 $I_m(u_h; \tilde{d}_1, \dots, \tilde{d}_r)$ をまだネットワークの入力ノードとして使っていない入力について計算し、その最大値を与える u_h を u_K とする。
4. ネットワークの入力ノードを 1 個増やし、そこに u_K を入力して学習を行う。得られたネットワークの出力を y_1, \dots, y_r とし、その誤差を $\tilde{d}_j = d_j - y_j$ とする。
5. 3へ戻る。

以上を停止基準を満たすまで行う。

最初にネットワークの入力として用いる u_H は、教師信号 d_j を説明する能力が最も高いものを選ぶが、二つめからは、教師信号 d_j のうちまだ説明できていない部分 $\tilde{d}_j = d_j - y_j$ についての説明能力に基づいて選ぶ。

重要性尺度 I の値が最も大きい入力ノードを逐次追加しながらネットワークを形成していくことによって、できるだけ入力ノード数の少ない一般性の高い NN を構築する。

ネットワークへの入力追加の停止基準は、 NN の構造決定でも一部用いられている AIC に類似の評価基準の最小化基準を用いる。

3.2 問題点

本来は選ばれるべき入力変数を k 個とすると、 $I_m(u_1, \dots, u_k; d_1, \dots, d_r)$ など、 k 個の入力変数を同時に考慮した尺度に基づいて入力変数を選定すべきである。しかし、相互情報量の計算には入力変数が各値をとる確率の推定が必要であり、入力変数の組を同時に条件とした確率を正しく求めるには実現不可能なほど多数のデータが必要となる。

そのため、ある入力ノードを用いたネットワークにおける誤差 \tilde{d} 、つまりその入力ノードだけでは教師信号を説明できなかった部分について、これを最も良く説明できる入力ノードを選ぶという近似手法が用いられている。しかし、この近似では良い結果が得られない場合がある。

この方法のイメージは Fig.1 のようになる。相互情報量によって選ばれる u_j は考えられるあらゆる方法を使って \tilde{d} を最もよく説明するものとして選ばれる。これは u_i と y の

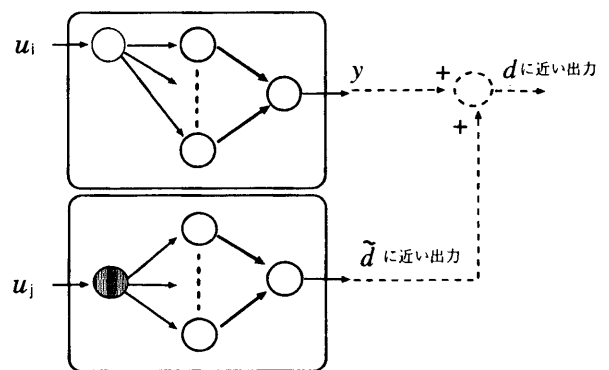


Fig.1 Network with a new input by (2)

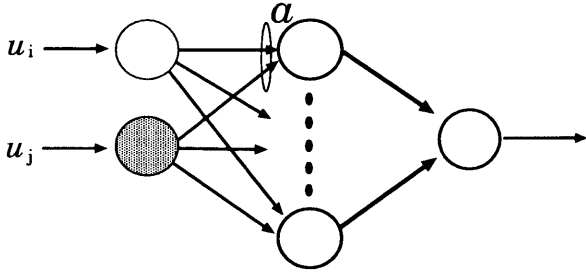


Fig.2 Network with a new input

間のNNとは無関係にあらゆるNNを試すことに相当する。しかし、実際はFig.2のように同じNNに u_j の入力ノードを加えるため、元のNNと共通部分がありこの部分は勝手に自由には変えられない。この点の考慮が必要である。

4. 今回提案する方法

4.1 方針

簡単に説明するため、3層のネットワークで1出力の時に2つめの入力を選定する場合を考える。 u_1 だけのNNに u_2 を加えた時のNNの出力、および中間層の入、出力を示すと

$$\begin{array}{ll} \text{\underline{u}_1だけのNN} & \text{\underline{u}_1とu_2のNN} \\ y = f(w_{i0}^1 + \sum_i w_{i1}^1 x_i) & \hat{y} = f(\hat{w}_{i0}^1 + \sum_i \hat{w}_{i1}^1 \hat{x}_i) \\ x_i = f(a_i) & \hat{x}_i = f(\hat{a}_i) \\ a_i = w_{i0}^0 + w_{i1}^0 u_1 & \hat{a}_i = \hat{w}_{i0}^0 + \hat{w}_{i1}^0 u_1 + \hat{w}_{i2}^0 u_2 \end{array}$$

u_1, u_2 : 初めに選んだ入力、及び次に選ばれる入力

y : 入力 u_1 で形成されたNNの出力

x_i : 第 i 中間層ノードの出力 ($i = 1, 2, \dots, n$)

a_i : 第 i 中間層ノードへの入力

w_{ji}^l : 第 l 層の第 i ノードから

第 $l+1$ 層の第 j ノードへの結合荷重

\hat{y} : 入力 u_1 と u_2 で形成されたNNの出力

\hat{x}_i : 入力 u_1 と u_2 で形成されたNNの中間層出力

\hat{a}_i : 入力 u_1 と u_2 で形成されたNNの中間層の入力

\hat{w}_{ji}^l : 入力 u_1 と u_2 で形成されたNNの第 l 層の第 i ノードから第 $l+1$ 層の第 j ノードへの結合荷重

f : シグモイド関数

2つのNNを比較すると結合荷重の値以外に異なるのは a_i と \hat{a}_i の中身である。そこで、教師信号に相当する \hat{a}_i の希望値の近似値 a_i^o を求め、これを入力の選定に利用する。本来、NN出力が教師信号に近づくような入力を選ぶべきであるが、Fig.2に示したように入力を付加する前後で中間層から出力に変る部分のNN構造が不変であること

を考慮に入れるため、中間層ノードにおける a_i の値を利用する。こうすることにより \hat{a}_i と u_2 の関係は線形であるので、大量のデータを必要とする相互情報量を用いなくとも線形理論の枠組内で最も良い u_2 を選ぶことができる。したがって、線形理論の枠組内で、 $I_m(u_h; \hat{d}_1, \dots, \hat{d}_r)$ の代わりとなる新しい重要性尺度 I を提案する。

4.2 重要性尺度

\hat{a}_i がその希望値に近くなり、従ってNN出力 y が教師信号 d に近くなる u を重要な u であるとする。そこで \hat{a}_i の希望値が必要である。希望値を直接知ることはできないので、その近似値 a_i^o を求める必要がある。

ここでは、一次近似を用いて a_i^o を求める。 a_i が $a_i^o = a_i + \Delta a_i$ に変化したとき y が $d = y + \Delta y$ になったとする

$$\begin{aligned} d &= y + \Delta y \\ &\simeq y + \sum_i \frac{\partial y}{\partial x_i} \Delta x_i \\ &\simeq y + \sum_i \frac{\partial y}{\partial x_i} \frac{\partial x_i}{\partial a_i} \Delta a_i \\ d - y &\simeq \sum_i \frac{\partial y}{\partial x_i} \frac{\partial x_i}{\partial a_i} \Delta a_i \end{aligned} \quad (3)$$

a_i の希望値とのずれ $\Delta a_i = a_i^o - a_i$ を、既に選んだ u_1 と新しい入力候補 u とを組み合わせ、近似表現する。

$$\Delta a_i \simeq \alpha_{i0} + \alpha_{i1} u_1 + \alpha_{i2} u_2 \quad (4)$$

(4)を(3)に代入すると

$$d - y \simeq \sum_i \frac{\partial y}{\partial x_i} \frac{\partial x_i}{\partial a_i} (\alpha_{i0} + \alpha_{i1} u_1 + \alpha_{i2} u_2) \quad (5)$$

となる。これから、係数 $\alpha_{i0}, \alpha_{i1}, \alpha_{i2}$ を(5)の両辺の間の差が最小となるよう最小2乗法で求める。

最小2乗法で求めた係数を用いて計算した $d - y$ の近似値と実際の $d - y$ が似ているかどうかで入力情報の重要性を判定できる。入力の組と $d - y$ の重相関係数の2乗を重要性尺度 $I_T(u)$ とする。

$$R^2(u_1, u, d - y) = \frac{\sum \{(e - \bar{e})(\hat{e} - \bar{\hat{e}})\}^2}{\sum (e - \bar{e})^2 \cdot \sum (\hat{e} - \bar{\hat{e}})^2} \quad (6)$$

ただし $e = d - y$ 、 \hat{e} は(5)の右辺の係数として最小2乗法で求めた値を使ったときの $d - y$ の近似値である。 $\bar{e}, \bar{\hat{e}}$ は e, \hat{e} の平均値、 \sum は全学習用サンプルについての和を表す。重相関係数 R は $-1 \leq R \leq 1$ の値をとり、 $|R|$ が1に近いほど e と \hat{e} が似ていることを表す。

$I_T(u)$ は入力を u_1, u_2, \dots と追加する度に $R^2(u_1, u_2, u, d - y), R^2(u_1, u_2, u_3, u, d - y), \dots$ のように計算していく。従来の方法では、既に選んだ入力のこととは考えていなかった。これはデータ数の制約のためで

あったが、線形の枠組で考えると既に選んだ入力との組にして近似誤差を最も小さくできる入力を逐次的に選ぶことにより、その制約が緩くなる。

4.3 他の重要性尺度

重要性尺度 $I(u)$ としては NN の重みや a_i^o の推定時の仮定の有無、近似の程度、計算量の大小が異なる他のものも考えることができる^{6),7)}。比較検討のために他の尺度を簡単に示す。

a_i^o の推定で、本来は $\Delta a_1, \Delta a_2, \dots$ を同時に考慮しなければならないが、個別に考える。(3)式で $\Delta a_1, \dots, \Delta a_{i-1}, \Delta a_{i+1}, \dots \equiv 0$ とすると

$$d - y \simeq \frac{\partial y}{\partial x_i} \frac{\partial x_i}{\partial a_i} \Delta a_i$$

よって

$$\left. \begin{aligned} \Delta a_i &\simeq \frac{d - y}{\frac{\partial y}{\partial x_i} \frac{\partial x_i}{\partial a_i}} \\ a_i^o &= a_i + \Delta a_i \end{aligned} \right\} \quad (7)$$

となる。中身を書き下すと

$$\Delta a_i \simeq \frac{d - y}{f'(w_{10}^1 + \sum_i w_{1i}^1 x_i) \cdot w_{1i}^1 \cdot f'(a_i)} \quad (8)$$

この値をもとに従来と同様に相互情報量を用いたものを重要性尺度 I_M とする。

$$I_M(u_h) = \sum_i I_m(u_h; \Delta a_{i1}, \dots, \Delta a_{ir}) \quad (9)$$

この尺度では $MI(u, \Delta a_i) = MI(u, (d - y) / \frac{\partial y}{\partial x_i} \frac{\partial x_i}{\partial a_i})$ に基づいている。従来との違いは $\frac{1}{\frac{\partial y}{\partial x_i} \frac{\partial x_i}{\partial a_i}}$ にある。これは、入出力の間の非線形変換の影響分を表している、これを考慮している分だけ近似の精度向上が期待される。

次に Δa_i を用いた線形理論の枠組内の尺度を示す。

u_2 を NN に加えて $a_i = a_i^o$ となるようにしたい。この時、重み w_{i0}^0, w_{i1}^0 を変えない($\hat{w}_{i0}^0 = w_{i0}^0, \hat{w}_{i1}^0 = w_{i1}^0$)と近似して $a_i = a_i^o$ と置く

$$\begin{aligned} a_i^o &= a_i + \hat{w}_{i2}^0 u_2 \\ a_i^o - a_i &= \Delta a_i = \hat{w}_{i2}^0 u_2 \end{aligned}$$

従って Δa_i と最も良く似た値を示す u を選べば良い。これは相関係数を使って選ぶことができる。

$$r(u, \Delta a_i) = \frac{\sum \{(u - \bar{u})(\Delta a_i - \overline{\Delta a_i})\}}{\sqrt{\sum (u - \bar{u})^2 \cdot \sum (\Delta a_i - \overline{\Delta a_i})^2}} \quad (10)$$

そこで重要性尺度 I_r を以下のように定義する。

$$I_r(u) = \sum_i [\{r(u, \Delta a_i)\}^2] \quad (11)$$

5. 入出力感度による削除的方法

生成的方法では基本的には $d - y$ を説明できる入力を追加してきた。しかし、教師データに対する表現能力を向上させるために必要以上に大きなネットワークとなる場合がある。

そこで、生成的方法によってネットワーク構成終了後、削除的方法を用いて不要な入力を削除する。不要な入力ノードを求める尺度として、入出力感度(入力の変動と出力の変動の比)を重要性の尺度とする方法が提案されている^{3)~5)}。生成的方法によって追加された入力をもつネットワークを出発点として、感度を重要性尺度とする削除的方法によって不要な入力を逐次取り除いていく。

以下のように記号を定め、多層(L層)型ニューラルネットワークにおける感度の計算を行う。

p : 第 p 番目の学習データ

$x_i^l(p)$: 第 l 層の第 i ノードの出力

w_{ji}^l : 第 l 層の第 i ノードから

第 $l + 1$ 層の第 j ノードへの結合荷重

$a_i^l(p)$: 第 l 層の第 i ノードへの入力

求めるべきは出力 $y_j = x_j^L(p)$ が入力 $u_i = x_i^0(p)$ の変動によって、どれだけ変動するかである。すなわち、 $\frac{dx_j^L(p)}{dx_i^0(p)}$ の値である。そこで、 $S_{ji}^l = \frac{dx_j^L(p)}{dx_i^0(p)}$ と定義すると、

$$S_{ji}^l(p) = \sum_k w_{ki}^l f'(a_k^{l+1}(p)) S_{jk}^{l+1}(p) \quad (12)$$

$$S_{ji}^l(p) = \frac{dx_j^L(p)}{dx_i^L(p)} = \delta_{ij} \begin{cases} 1 (i = j \text{の時}) \\ 0 (i \neq j \text{の時}) \end{cases} \quad (13)$$

以上の計算により、入出力感度 $S_{ji}^0(p)$ を求める。この入出力感度をすべてのデータについて計算する。

入力層ノード1つ1つにつき出力層ノードの数だけ計算した感度のうち、最も大きな値を持つものをその入力層ノードの出力層ノードに対する重要性の尺度として用いる。すなわち、入力層ノード i の重要性尺度は

$$I_i = \max_j \sqrt{\frac{1}{p} \sum_{p=1}^P \{S_{ji}^0(p)\}^2} \quad (14)$$

で与えられる。

それぞれの入力ノードで I_i を計算し、この値が小さいほど出力に対する影響力が小さいと考えられる。最小の感度をもつ入力ノードを1つ削除し、 NN の結合荷重を再学習する。停止基準を満たすまで感度計算、最小の感度

をもつノードの削除，結合荷重の再学習を繰り返す。

停止基準は生成的方法と同じくAICに類似の評価を用いる。

6. 例 題

6.1 対象および条件

次の関数から生成されるデータを，ニューラルネットワークによって補間近似させる。

$$d = x_1^2 + (x_1 + |x_1|)\{x_2 + (x_2 + |x_2|)x_3\} \quad (15)$$

変数 x_1 は常に d に影響を与えるが， x_2 は $x_1 > 0$ のときにのみ d に寄与し，さらに x_3 は $x_1 > 0$ かつ $x_2 > 0$ のときにのみ意味を持つ。 x_3 より x_2 ， x_2 より x_1 が入力情報として重要といえる。

入力の要，不要を適切に判断できるかを検証するため，偽の入力として x_1, x_2, x_3 にある確率で誤情報が入っている入力，全ての情報に平均値が0であるようなノイズを加えた入力，さらに無関係の乱数を用意した。

入力 u_i を以下に示す。

$$u_1 \sim u_3 : x_1 \sim x_3$$

$$u_4 \sim u_6 : x_1 \sim x_3 \text{の値を4割の確率でランダムに変える(値} -1, -0.9, -0.8, \dots, 0.9, 1)$$

$$u_7 : -0.5 \sim 0.5 \text{の乱数}$$

$$u_8 \sim u_{10} : x_1 \sim x_3 \text{に平均値0のノイズを含む}$$

各 x_i が $-1, -0.5, 0, 0.5, 1$ の値をとる計 $5^3 = 125$ 通りについての，10入力および d の値の組を学習データとして用いた。

検証データとして，学習データとは異なる値，各 x_i が $-0.75, -0.25, 0.25, 0.75$ の計 $4^3 = 64$ 通りを与えた。

入力追加，削除の停止基準として(16)式で表されるAICに類似の評価を用いる。評価 C がノードを追加する前より追加した後のほうが大きくなるまで入力を生成的方法を用いて追加する。 C が大きくなったということはノードを追加しないほうが良かったということの意味するが，今追加したノードそのまま削除するのではなく，削除的方法を行う。削除したネットワークの C が小さければさらに削除を続ける。最終的な入力ノードと結合荷重は C が最小の時のネットワークとする。

$$C = 2N \log(E) + \alpha K \quad (16)$$

E : 出力と教師信号の平均2乗誤差の和

N : 全データ数(学習データ + 検証データ)

K : ニューラルネットワークに含まれるパラメータ数

α : 調整係数

6.2 結果

評価 C の調整係数 $\alpha = 6.0$ としてデータの乱数系列が異なる(case 1~5)シミュレーションを5回行った。Table-1

に入力を追加，削除してネットワークを形成した際の選択された入力ノードの番号，評価 C ，及び検証データに対する平均2乗誤差を示す。今回提案した重要性尺度 I_T が妥当であることが確認できる。相互情報量を用いた従来の尺度 I_m や I_M では2つめの入力を選ぶ段階で不要と考えられる入力を選ばれ， C が大きくなりそれ以上入力は追加されなかった。NNの構造を考慮した分 I_M は向上が期待されたが，その効果は殆んどなかった。これは相互情報量が確率の推定値を必要とし，データ数が少ない場合にはその確率が正しく推定できないことが原因と考えられる。 I_T では確率を必要としないので少ないデータ数でも機能できている。また， I_T と比較することで既に選んだ入力と組にして次の入力を選ぶ効果が確認できる。case1, case2を見れば3つめの入力が既に選んだ入力と組み合わせ，正しく選ばれているのが確認できる。 I_T では構造を考慮したことによって，線形理論の枠組内で考えることができるため，確率の推定値を必要としない相関係数を尺度として用いることができた。また，既に選んだ入力と合わせて適切な入力を選ぶことが可能となった。この2点から少ないデータ数でも重要性尺度として機能することが確認できた。

次に学習データと検証データに対してのネットワークの平均2乗誤差の5回のシミュレーションの平均値を計算した。結果をTable-2に示す。入力を選定し，必要な入力のみでネットワークを構成することにより，学習データに対する誤差は増加しているが，検証データに対する誤差は小さくなっており，ネットワークの汎化能力が向上したことを確認できる。

Table-2 Average mean square error

Data set	NN with all inputs	NN with inputs selected by (6)
Training set	1.00×10^{-5}	5.07×10^{-5}
Validation set	5.66×10^{-4}	3.04×10^{-4}

7. ま と め

本研究では，階層型ニューラルネットワークへの適切な入力の選定を行うことを目的とした。具体的には入力情報の有用性を測る重要性尺度として，ネットワーク構造と既に選んだ入力を考慮にいたした生成的方法での尺度を提案した。例題から提案した重要性尺度によって，入力の要不要の適切な判断と汎化能力の向上が実現できることを確認した。

参 考 文 献

- 1) 村田; 藤井; 池田; 平澤; 相良: 計測自動制御学会論文集 **31-2**(1995)236/243.
- 2) Gelbwachs, J. A.; Klein, C. F.; Wessel, J. E.: *IEEE Trans. on Neural Networks*. **5-4**(1994)537.

Table-1 Results of input variable selection. E_v is mean square error on validation set. Letter 'a' after a node number means that the node is added by the constructive algorithm, while 'd' indicates that the node is deleted by the destructive algorithm. Underlined figures are final values and indicates the node numbers eventually selected by the method.

case	I_m			I_M			I_r			I_T			
	Node No.	C	E_v $\times 10^{-4}$	Node No.	C	E_v $\times 10^{-4}$	Node No.	C	E_v $\times 10^{-4}$	Node No.	C	E_v $\times 10^{-4}$	
1	1 a	-861.8	19.6	1 a	-861.8	19.6	1 a	-861.8	19.6	1 a	-861.8	19.6	
	4 a	-805.6	21.4	8 a	-794.6	23.7	2 a	-1059.3	7.9	2 a	-1059.3	7.9	
	4 d	-861.8	19.6	1 d	-874.4	20.4	10 a	-2199.0	1.7	3 a	-2323.5	1.2	
							9 a	-2212.2	1.4	6 a	-2419.3	0.8	
							5 a	-2172.3	1.4	8 a	-2262.3	1.3	
							5 d	-2233.7	1.5	6 d	-2353.6	0.9	
							9 d	<u>-2343.0</u>	<u>1.3</u>	8 d	<u>-2452.3</u>	<u>0.8</u>	
							10 d	-1060.6	7.9	3 d	-1061.3	7.7	
	2	1 a	-861.8	19.6	1 a	-861.8	19.6	1 a	-861.8	19.6	1 a	-861.8	19.6
		8 a	-819.5	25.4	8 a	-817.2	26.3	9 a	-1155.6	15.7	9 a	-1155.6	15.7
1 d		-896.9	20.1	1 d	-899.1	20.1	8 a	-1201.9	15.2	10 a	-1968.1	2.2	
							2 a	<u>-1222.7</u>	<u>15.5</u>	6 a	-1887.8	3.5	
							4 a	-969.3	43.3	6 d	<u>-2003.0</u>	<u>2.3</u>	
							4 d	-963.8	48.6	10 d	-1278.7	12.0	
3		1 a	-888.7	19.8	1 a	-888.7	19.8	1 a	-888.7	19.8	1 a	-888.7	19.8
	8 a	-819.9	21.2	8 a	-819.5	20.9	2 a	-1082.0	9.1	2 a	-1082.0	9.1	
	8 d	-888.0	20.0	8 d	-888.8	19.8	3 a	<u>-1676.8</u>	<u>10.4</u>	3 a	<u>-1676.8</u>	<u>10.4</u>	
							8 a	-1219.4	30.3	9 a	-1662.7	9.2	
							8 d	-1486.4	18.1	9 d	-1519.7	16.4	
4	1 a	-864.0	19.6	1 a	-864.0	19.6	1 a	-864.0	19.6	1 a	-864.0	19.6	
	4 a	-855.1	39.7	8 a	-842.6	22.0	2 a	-1063.8	8.0	2 a	-1063.8	8.0	
	1 d	-836.8	24.2	8 d	-864.2	19.6	3 a	-2435.1	1.1	3 a	-2435.1	1.1	
							9 a	-2519.2	0.7	10 a	-2443.0	0.9	
							5 a	-2529.3	0.7	5 a	-2338.5	1.1	
							4 a	-2521.9	0.6	5 d	-2425.6	0.9	
							4 d	<u>-2560.2</u>	<u>0.6</u>	10 d	<u>-2589.6</u>	<u>0.7</u>	
							5 d	-2546.7	0.7	3 d	-1065.2	7.8	
	5	1 a	-861.8	19.6	1 a	-861.8	19.6	1 a	-861.8	19.6	1 a	-861.8	19.6
4 a		-796.5	29.3	8 a	-643.9	99.3	2 a	-1059.4	7.9	2 a	-1059.4	7.9	
4 d		-861.8	19.6	8 d	-859.8	20.4	3 a	-2306.3	1.0	10 a	-2145.9	1.5	
							9 a	-2298.6	0.9	3 a	-2288.7	1.2	
							9 d	<u>-2364.4</u>	<u>1.0</u>	4 a	-2247.5	1.4	
							3 d	-1058.0	8.0	4 d	-2299.9	1.2	
										10 d	<u>-2356.8</u>	<u>1.0</u>	
										3 d	-1060.9	7.8	

3) 阿部: 九州大学工学部電気工学科卒業論文 (1995).

4) 野田; 村田; 平澤: 計測自動制御学会九州支部学術講演会 (1995)221.

5) 野田; 村田; 平澤; 阿部: 電気学会電子・情報・システム部門大会予稿集 (1996)531.

6) 中園; 村田; 平澤: 電気学会電子・情報・システム部門大会予稿集 (1997)625.

7) 中園; 村田; 平澤: インテリジェント・システム・シンポジウム予稿集 (1997)277.