

## ニューラルネットの音声・画像・制御への応用

高木, 英行  
松下電器産業株式会社中央研究所情報科学研究室

<https://hdl.handle.net/2324/4481537>

---

出版情報 : pp. 1-43, 1988-09-28. マイクロエレクトロニクス研究開発機構  
バージョン :  
権利関係 :

## I. ニューラルネットの音声・画像・制御への応用

松下電器産業株式会社 高木英行  
中央研究所情報科学研究室

### 要 約

ニューラルネットの3分野への応用について述べる。第1は音声情報処理への応用で，ホルトマン抽出・音声合成のための韻律性の制御・音韻セグメンテーション，第2は不特定筆者の手書き英数文字認識，第3はニューラルネット駆動型ファジィ推論である。これらの実験結果を概観し今後の研究取組に対する展望を述べる。

Artificial neural network to three application fields is described in this paper. First application field is processing, and it is applied to the formant extracton, prosody control for text to speech, and phonetic segmentation. Second application field is character recognition, and it is applied to the recognition of hand-written alphanumeric characters. Final application field is control, and the combination of fuzzy reasoning and neural network that is called artificial\_neural\_network-driven fuzzy reasoning is proposed. Each experiments are described and through these experiments, research of neural network from now is considered.

### 1. はじめに

昨今のニューラルネットへの関心の高まりは以前の超電導のような感を呈し，研究者のみならず話題を多く提供してきたが，最近はやうやく落ち着きを取り戻しつつある。ブームに対する批判もあったが，研究者数の急激な増加によって特に応用面での進歩が見られた。

ところで，ニューラルネットの工学的研究の方向には3種類あるように見受けられる。第1は，生体の神経情報処理メカニズムを極めようとするアプローチで，生理学的知見をできる限り取り入れ，未解明な点は大胆な仮定を立てて工学モデルを構築しようとするものである。生体計測のし易さから視覚系の，

特に初期視覚のモデル化が進んでおり、ハードウェア化もされつつある<sup>1)</sup>。

第2は、脳・神経情報処理のアルゴリズムを求める、神経計算学とでもいう方法である<sup>2)</sup>。従来のコンピュータを支えたソフトウェア・アーキテクチャがあるように、脳・神経のソフトウェア・アーキテクチャも存在するに違いない。それは、自ら学び素早く的確に（正確などとは言えないまでも）判断する我々の脳の存在がこれを証明している。これを解明する過程で得られた成果が実際の脳と同じであろうと違おうと、工学的には大変有用であることは想像に難くない。

第3は神経とは全く別の次元でツールとしてニューラルネットを用いる立場である。多くの工学研究者はこの部類に属する。例えば、パターン認識の研究者は認識率さえ上がるならば、どのような方法でも試してみるであろう。ニューラルネットはそのうちの1つのツールに過ぎないわけである。勿論、その性能はフルに発揮させたいわけで、そのためには何をどの様に加工して入力し何を取り出すか、つまり目的とするタスクに応じていかなるニューラルネットのモデルを作るかが大きな研究課題になる。また、第1・第2の研究成果もこの目的に合致するものはどれでも利用する。

筆者等は音声認識への前処理機構として、第1の立場である聴覚末梢系の工学モデル化にも取り組んでいるが、他の研究と平行しながらじっくり時間を掛ける必要があろうと考えている。そこで、本解説では、第3の応用に主眼を置いた筆者の最近の研究を紹介し、今後への取組を展望する。なお、ここで紹介する研究は、既に発表済みまたは投稿済みの内容であり、詳細は参考文献を御覧いただきたい。

## 2. ニューラルネットの導入説明

脳神経系での情報処理機構を数理工学的に解明しようとの試みは、連想記憶や学習機械の研究として古くから行なわれていた<sup>25, 2)</sup>。一時は学習機械パーセプトロンに多くの研究者の関心が集まり盛んに研究されたが、Winsky等の論文<sup>26)</sup>ですっかり熱が冷めてしまい、以後地道な研究期間が続いた。近年多層パーセプトロンの効果的な学習アルゴリズム backpropagation<sup>4)</sup>が発表され、ま

た、巡回セールスマン問題<sup>27)</sup>やNETtalk<sup>7)</sup>のようなデモ効果の大きい応用研究が発表されて以来、第2次ニューラルブームが起きてきたのはご承知の通りである。

本解説ではこれらの一般解説をすることは目的ではないので、ニューロンモデルの導入説明後は、ここで取り上げるニューラルネットに限定して定義を簡単に行う。

Fig.1 は生物のニューロンである。ニューロンは他の複数のニューロンからシ

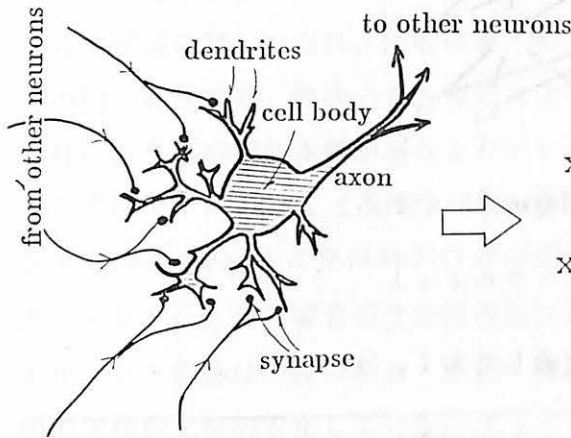


Fig.1 ニューロン

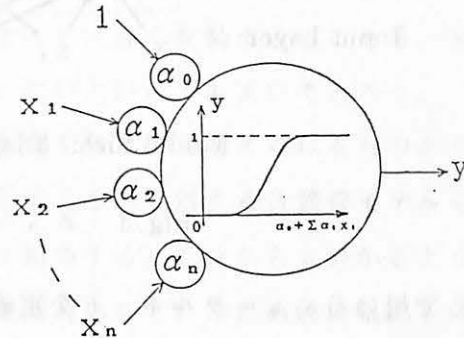


Fig.2 ニューロンモデル

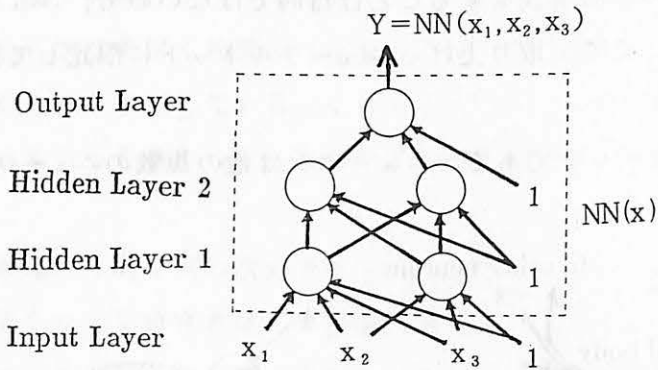
ナプスを介して信号を受け取る。これ等の入力信号が閾値以上であれば神経パルスを発火し、以下であれば発火しない。以上の生理学的知見を工学的に表現したモデル例がFig. 2 に示す多入力-出力の非線形回路であり、数理モデルとして入力  $x$  と出力  $y$  の関係を(1)式で表わす。

$$y_i = f(\sum \alpha_{ji} x_{ji} + \alpha_0) \quad (1)$$

$$f(z) = \frac{1}{1 + \exp(-z)} \quad (2)$$

ここで  $\alpha$  はシナプス結合をモデルに反映させた結合強度である。(2)式に sigmoid 関数を用いるのは、2 値関数的でかつ backpropagation で要求される微分可能性を満足しているからである。

このニューロンモデルをネットワーク結合したものがニューラルネットである。ニューラルネットには多くのモデルがあるが、本解説ではFig. 3に示すfeed-forward型のみを用い学習アルゴリズムはbackpropagationを用いる。



Model Size: 3 layer [  $3 \times 2 \times 2 \times 1$  ]

Fig.3 ニューラルネット

ここで用いるニューラルネットを定義しておく。

[本解説中で用いるニューラルネットの定義]

- (a)  $k$  層パーセプトロンとは 1 層の出力量と  $(k - 1)$  層の中間層を持つモデルである。
- (b) 神経回路網の入出力関係を(3)式で表現する。

$$y = NN(x) \quad (3)$$

- (c) モデルの規模を  $k$  層  $[U_0 \times U_1 \times \dots \times U_k]$  で表現する。  $U_i$  は各々入力層, 中間層, 出力層の神経細胞モデルの数である。
- (d) 入力層と中間層は, 上記(c)以外に定数 1 のユニットをもつ。このユニットは下層の神経細胞モデルからの結合はないが, 上層の神経細胞とは結合を持つ。
- (e) 隣接する層の神経細胞モデルは全て結合する。同一層内や層を飛び超えての結合はない。

(d)は(1)式の $\alpha_0$ を表現するためであるが、3.1節ホルマント抽出と3.2節プロソディ制御及び4.1節英字認識で用いたモデルにはまだ、この定数1が考慮されていなかった。特に以後では(c)の表現を用いてモデル構成を示すこととする。例えば、Fig.3は3層〔3×2×2×1〕と表現する。

### 3. 音声情報処理への応用

#### 3.1 ホルマント抽出<sup>3)</sup>

##### 3.1.1 背景と目的

音声認識の難しい点は、話者の個人性・音韻性の分離が未だできず、同一話者の同一音韻でも、前後の音韻環境によって大きく変動する点にある。言い代えれば、未だ音声の本質が掴みきれていないといってもよいであろう。

この状況下において、音韻性に最も影響を与えらるものがホルマントである。Fig.4に発声時の口腔の形状とこれに対する音響管モデルを示す。ホルマントはこの音響管の共振周波数に相当する。これからも分かるように、ホルマント軌跡は口腔の動き・形状に直接対応しているため連続軌跡を描き時間的変動が比較的安定している。従ってホルマントに着目すれば、音声の時間的な微小変動を吸収し、大局的な時間的変化を掴むことができる可能性があり、

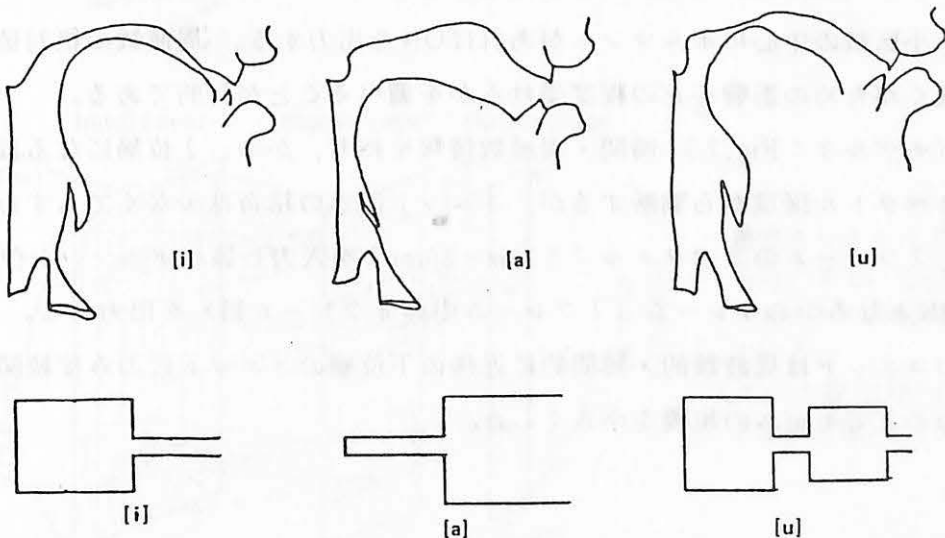


Fig.4 発声時の音響管モデル

音声認識性能の向上に大きく寄与するものと思われる。そこで筆者等は音声認識における変動要因吸収の一方法として、前述した聴覚末梢系の生理学的知見を基に前処理部を構築し、大局的なスペクトルの動きを追うことで大きな誤りを犯さない音声認識システムを考えている。この考えを実現するための特徴抽出として、今回はホルマント抽出を取り上げ、ニューラルネットで抽出フィルタリングを行う。ここでは聴覚神経系のモデル化は目的とせず、信号処理への応用の可能性を探ることを目的とする。本実験におけるホルマント抽出は画像処理と言った方が相応しいのかもしれない。

### 3.1.2 モデルの構成

スペクトルを入力しホルマントを出力する3種類のモデルを構成した。以下各モデルの狙いと構成を述べる。

(モデル1: Fig.5) 時間情報を含まないモデル。1フレーム分の対数パワースペクトル(5 kHzまで64点)を入力とし、該当ホルマント位置のみONとなる64点周波数スペクトルを出力する。フレーム間情報を含まないがためにフレーム毎のスペクトル変動の影響をどの程度受けるかを調べるのが目的である。

(モデル2: Fig.6) 時間・周波数情報を持つが、共に部分情報しか持たないモデル(468 Hz×20 ms)の小区域で全スペクトログラム上をスイープ入力し、小区域の中心にホルマントがあればONを出力する。周波数の絶対値情報を欠くがための影響をどの程度受けるかを調べるのが目的である。

(モデル3: Fig.7) 時間・周波数情報を持ち、かつ、上位層になるほど広いスペクトル区域から判断するが、ユニット同志の結合は少なくともすむモデル。7フレームのスペクトル(5 kHz×35ms)を入力し該当ホルマント位置のみONとなる中心フレーム(7フレーム中の4フレーム目)を出力する。上位層のユニットは周波数的・時間的に近傍の下位層のユニットにのみ接続関係を持つことでモデルの規模を小さくした。

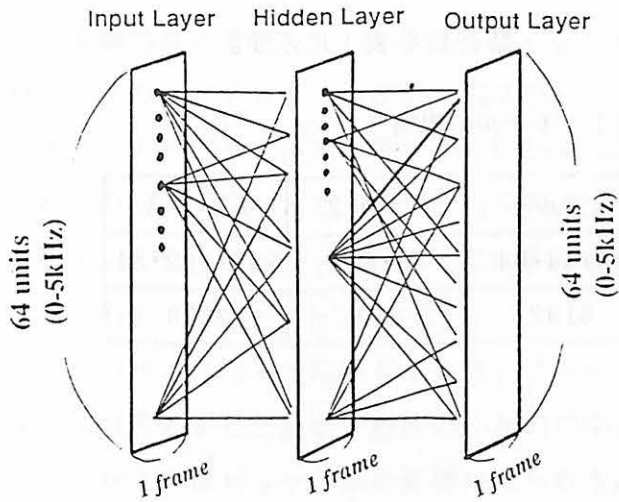


Fig.5 モデル1  
フレーム毎にスペクトルを入力する。時間情報は持たない。  
(ユニット数:64-64-64,  
結合数:8192)

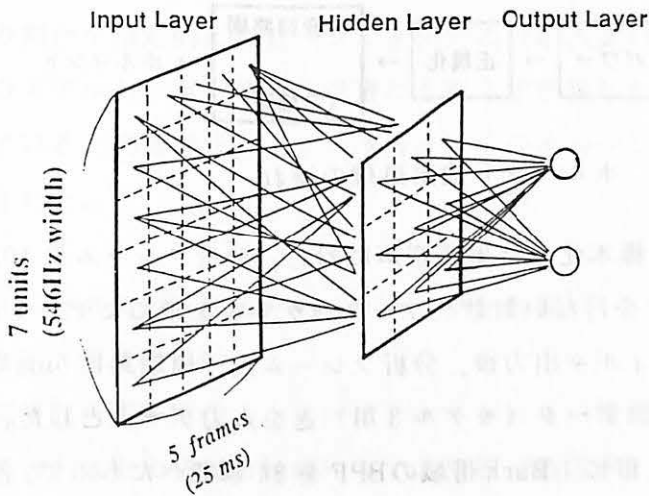


Fig.6 モデル2  
スペクトログラム上をスweepし部分的なスペクトルを入力する。  
(ユニット数:35-9-2,  
結合数:333)

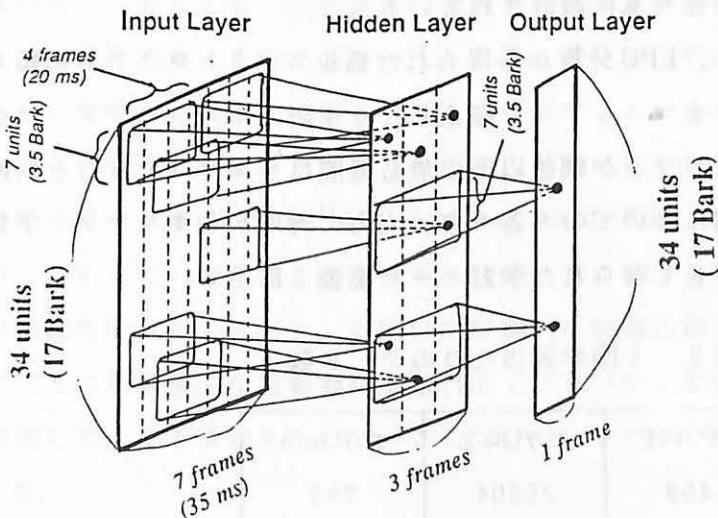


Fig.7 モデル3  
聴覚スペクトログラムを入力する。上位層のユニットのみに結合している。  
(ユニット数:238-102-23,  
結合数:3570)



各モデルのユニット数・ユニット結合数を表1に示す。

表1 モデルの規模

	モデル1	モデル2	モデル3
ユニット数	64・64・64	35・9・2	238・102・34
結合数	8192	333	3570

### 3.1.3 実験方法

実験の流れをFig8に示す。

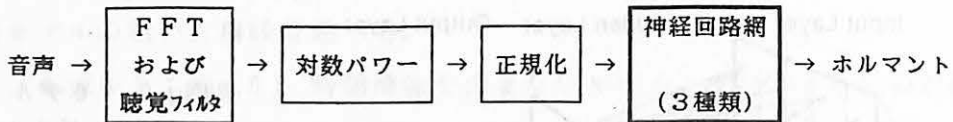


Fig.8 ホルマント抽出処理の流れ

(a) 入力データ 20kHzで標本化した単語音声に対し、分析フレーム長10ms・フレーム周期5msでDFTを行ない対数パワースペクトルを求めたデータ(モデル1と2用)と、聴覚フィルタ出力後、分析フレーム長・周期共に5msで対数パワースペクトルを求めたデータ(モデル3用)とを入力データとした。

聴覚フィルタは1/2 Bark毎に1 Bark帯域のBPFを34個並べたもので、各フィルタは基底膜の共振特性を基に設計されている<sup>5)</sup>。

(b) 学習データ 初めに、LPC分析から得られた極をスペクトログラムに照らし合わせ、誤抽出データをマニュアルで修正した6単語のホルマントデータを作成する。次に入力信号パワーが閾値以下の無音区間はホルマントなしと判断し、閾値以上の時間区間についてのみ各モデル入力に対応するホルマント学習データを作成する。こうして得られた学習データを表2に示す。

表2 1回学習当たりのデータ数

モデル1	モデル2	モデル3
1459	29504	785

### 3.1.4 実験結果と考察

学習データのスペクトログラム出力の例をFig.9 (a0)に示す。また、未学習データのスペクトログラム出力の例をFig.9 (b0)に示す。これら(a0) (b0)を入力した時のモデル1・モデル2の出力をFig.9 (a1) (a2)とFig.9 (b1) (b2)に示す。また、(a0) (b0)に対応する聴覚スペクトルを入力したモデル3の出力をFig.9 (a3) (b3)に示す。

モデル1の出力には不連続部が見られ、フレーム間情報の必要性が伺える。モデル2は対象とする周波数・時間の小領域が小さすぎるのではないかと思われる。モデル3にはピッチ周期の影響が見られる。このため聴覚フィルタ出力とFig.9 (a0), (b0)とを比較検討した結果、ホルマントに対するピッチの影響の割合が相対的に聴覚フィルタ出力の方が大きいためであると思われる。何れのモデルも、部分情報の学習にも拘らず学習した音韻のホルマントを抽出できていると判断でき、ニューラルネットのホルマント抽出に対する有効性が確認された。

## 3.2 音声規則合成のためのプロソディ制御<sup>6)</sup>

### 3.2.1 背景と目的

音声規則合成を行なうためにはテキストにない情報を付加しなければならない。テキストは主に音韻の種類に関する情報であるが、これとて時と場合による例外があり、ましてや、声の高さ、音の持続時間、強弱等、音声の自然さを表現する情報はテキスト中にはない。これ等、音声の自然さの情報をプロソディ（韻律）と呼ぶ。

音声合成へのニューラルネットの応用としては、英語のスペリングから発音記号を推定するNETtalk<sup>7)</sup>があるが、日本語の場合、文字と発音記号はほぼ一対一対応しているので、問題はそれ以降のプロソディ制御にある。しかも、この制御規則はピッチパターンを除いて経験的に制御規則が作られているのが現状で、この制御規則を自動獲得する目的でプロソディ制御にニューラルネットを適用した。まず手始めとして、ここではピッチパターン制御についての実験を述べる。

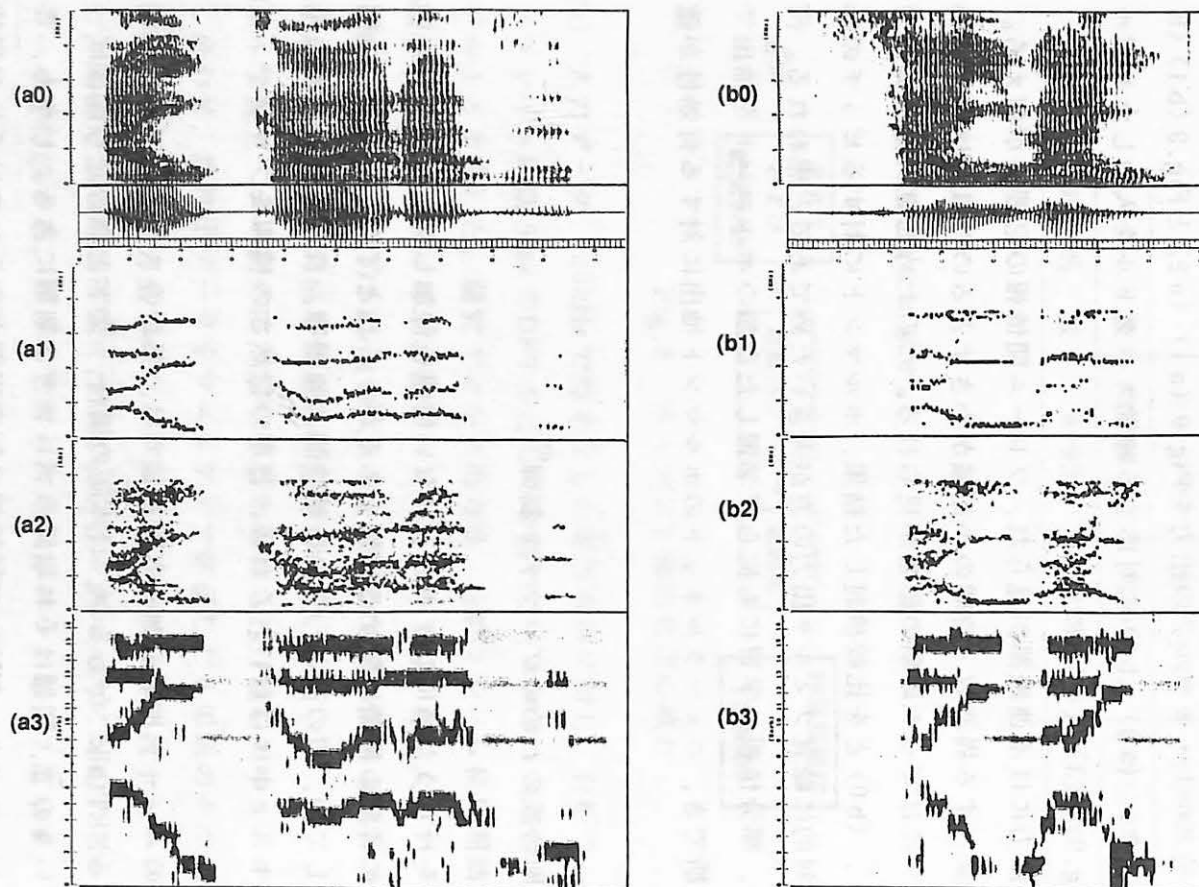


Fig.9 (a0)と(b0)は各々学習データと未学習データ  
(a1)～(a3)は(a0)を入力した時のモデル1～3の出力  
(b1)～(b3)は(b0)を入力した時のモデル1～3の出力

### 3.2.2 特長—方言性・個人性表現

ピッチパタンの制御について言えば、藤崎モデル<sup>8)</sup>で代表されるように発音記号からピッチパタンへの変換関係が陽になっていた。本方法ではニューラルネットを用いるので、この変換関係が回路網の接続強度として陰に表現される。その代わりに、数式表現された変換モデルより表現の自由度が高まるため、学習データに順応しやすいと考えられる。ましてや、他のプロソディ制御規則は確としたものがないだけに、本方式には

- (1) 方言性や個人性が容易に表現でき得る
- (2) 方言性や個人性の表現は専門家によるモデルが不要で、学習によって自動的に獲得される
- (3) 方言性や個人性の変更は神経回路網モデルの重み係数を入れ替えるだけである

の特長がある。

### 3.2.3 モデルの構成

音韻種とアクセント型を入力し、実測ピッチを出力するよう学習した。今回は簡単のため、最大7モーラまでの音声を対象とするモデルで行った。Fig.10にモデル構成を示す。Fig.10は1モーラ分の入力ユニットで、実際のモデルはこの7モーラ分のモデル規模となり、隣接する2層間はすべての結合がなされている。

(a) 入力層 子音の調音位置・母音の重視・モデルの規模・学習時のエラー収束の容易さ等を考慮して1モーラ当たり13ユニットをFig.10のように割り振った。入力音韻列が各入力ユニットに該当しない場合は0を、入力モーラ数以外のユニットには-1を与える。該当する場合は(0~1)の間を該当ユニットの受持音韻数で配分する。例えば、/nya/に対する場合は、拗音に対し第2ユニットが1/3、/n/に対し第3ユニットが6/8、母音に対し第8ユニットが1で残り10個のユニットは0として1モーラ分入力データを作る。

(b) 出力層 今回の実験はピッチ制御のみなので、音韻継続長で正規化した時のピッチパタンを出力するように設計した。具体的には、モーラ毎に子音・母



で修正した。次に 3. 2. 3 節の代表ピッチを求め、1 単語分を 1 つの学習データとした。

学習データの音韻記号を本方式のニューラルネット(264 回学習後)に入力して得られたピッチパターンを Fig.11 (a) に示す。また、未学習データの例を Fig.11 (b) に示す。ドット線は実測ピッチである。音韻区間長は実測値をそのまま用いた。なお、語頭や /k/ の推定ピッチが 50Hz 近傍であるのは、学習時に 50Hz としたためである。

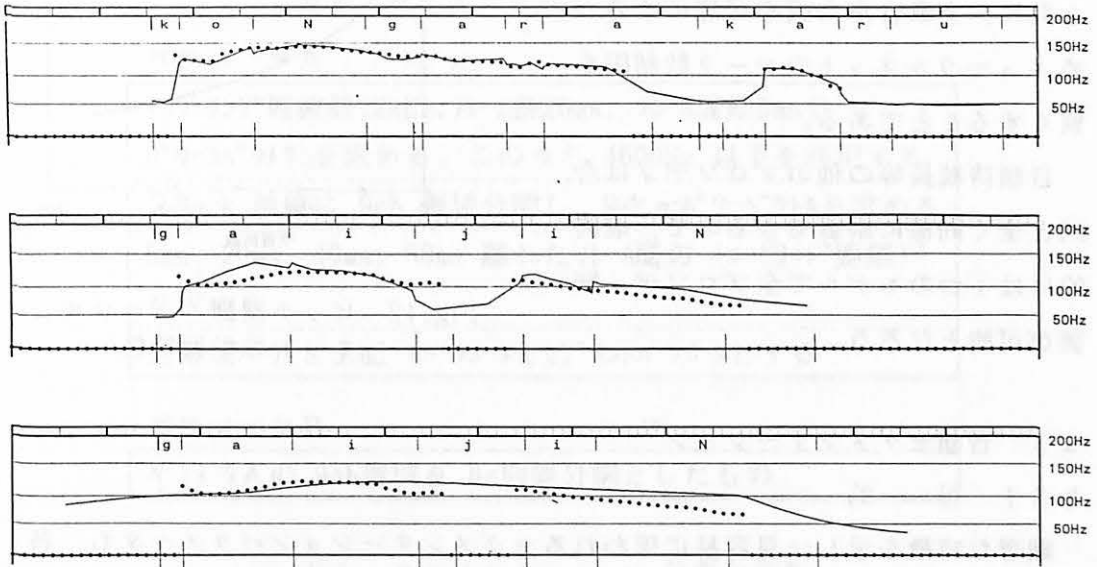


Fig.11 ピッチ推定例 … ドットは実測ピッチ  
 (a) 本方式(学習データ) (b) 本方式(未学習データ)  
 (c) 藤崎モデル  
 学習回数は 264 回

### 3. 2. 5 考 察

全体的に妥当な推定をしており、本方式の有効性が確認できた。語頭のピッチ下降や /j/ の推定に比較的誤差が多いのは、ピッチがない場合の影響が大きいためであろう。他にも有声子音でピッチなしと推定する例が見られたが、学習データに依存する問題であり、実測のピッチ抽出精度が上がれば解決するものと考えられる。比較のため、藤崎モデルで推定したピッチパターンを Fig.11 (b) に示す。

本実験で用いたニューラルネットの学習特性をFig.12に示す。この誤差収束曲線から、まだ性能の向上が期待される。

今回は学習した音韻環境が少ない(490単語)ので、未知入力に対してはFig.11ほど推定はできていない。今後の課題は、学習データ数を増やし各種の音韻環境に対応できるようにすると共に、文章音声合成を可能にするためニューラルネットのモーラ数制限を無くすることである。

音韻持続長等の他のプロソディは今回と全く同様に制御できるので、最終的には1つのモデルで全プロソディ制御が可能となろう。

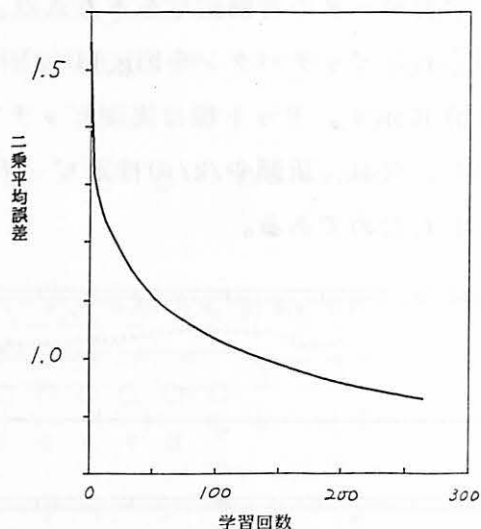


Fig.12 ピッチ制御ニューラルネットの学習特性

### 3.3 音韻セグメンテーション<sup>9)</sup>

#### 3.3.1 目的

顕著な特徴を示し一見容易に思われるセグメンテーションパラメータも、各種音韻環境に適用すると単純な閾値だけでは性能が頭打ちになり、例外処理や複雑な判定ルールが必要になる。この経験に基づいた部分をニューラルネットで代行しようとするのが本研究の狙いである。同様な研究に、約10モーラの2文章を対象としてニューラルネットの学習方法に主眼を置いた報告がある<sup>10)</sup>。ここでは全単音節を対象に、入力パラメータの選択・モデルの構成・発声時期の違いや話者の違いに対するロバスト性を評価することで、連続音声への適用に向けた知見を得ることを目的とする。ただし、ここでの報告は実験の中間段階である。

#### 3.3.2 実験条件とニューラルネットモデル

対象とするセグメンテーションの種類は表5に示した詳細なものとする。表3に示した音声資料から表4のパラメータAとBを求める。各パラメータは最大値と無音区間の平均との間を線形に正規化してニューラルネットへ入力する。

表3 音声資料

SM0	男性話者A発声 101単音節	(学習用)
SM1	同上 (SM0との発声間隔は1年半)	(評価用)
SM2	男性話者B発声 101単音節	(評価用)
SF1	女性話者C発声 101単音節	(評価用)
WM1	男性話者A発声単語	(評価用)

表4 セグメンテーションパラメータ

パラメータA

サンプリング周波数12kHz, フレーム長20ms, フレーム周期5msで パワースペクトラムを求める。このうち 4600Hz 以下を利用する 2Bark 単位に 9ch 帯域分割し、9次元パワーベクトルを求める 5ms, 20ms, 40ms, 60ms 離れたフレーム間の (ユークリッド距離) <sup>2</sup> を求める 全帯域パワーと上記 4パラメータをセグメントパラメータとする
---

パラメータB

パラメータAの 9ch帯域を Hz均等分割としたもの
----------------------------

表5 セグメンテーション位置の種類

(1) 語頭	(5) C・V境界
(2) クラスタ・C境界	(6) V・出渡り境界
(3) C・Y境界	(7) 語尾
(4) Y・V境界	

Yは拗音・/y/、Cはその他の子音、Vは5母音と濁音を示す

矛盾した学習を防ぐため、セグメンテーション位置に隣接するフレームは学習からはずした。これは、5ms程度のズレではニューラルネットへの入力パターンは殆ど変わらないと思われるにも拘らず、学習時にセグメンテーション位置の有無で全く違う学習を行ってしまうからである。



入力層ユニット数は（入力フレーム数×セグメンテーションパラメータ数）である。学習の効率化のため、(1) 実際のニューラルネットではこれ以外に、上位層とのみ結合する定数1のユニット1個を入力層と中間層に持たせた、(2) チェビシェフ加速法<sup>11)</sup>をヒントに、重み係数修正量 $\Delta w(t)$ の計算に $\Delta w(t-1)$ だけでなく $\Delta w(t-2)$ も利用した。

### 3.3.3 実験

#### (a) セグメンテーション位置の細分化の効果

同じセグメンテーションでも語頭と語中ではニューラルネットへの入力パターンは大きく異なるため、この違いがニューラルネットに矛盾した学習をさせて収束や性能に悪影響を与えることが考えられる。そこで、セグメンテーション位置の種類を分類して違う出力層ユニットに割り当てるか、1つの出力層ユニットのON/OFFで判定するか、の比較実験によってこの影響の有無を調べる。

表5のセグメンテーション位置7種類とセグメンテーション位置でないこと<sup>1)</sup>の計8種を出力層ユニットに対応させた2層〔T×5, 50, 8〕〔構成のモデル1と、セグメンテーション位置の有無を1つの出力層ユニットに対応させた2層〔T×5, 50, 1〕構成のモデル2とを比較する。Tはニューラルネットへ一度に入力するフレーム数を示し、時間換算で30ms, 50ms, 70ms, 90ms, 110ms, 130ms, 150msの7種類である。用いる音声資料は学習時と同じくSM0のパラメータAに対してである。

#### (b) open実験

実験(a)を音声資料SM1, SM2, SF1について行なう。

#### (c) 帯域分割の違いによる性能比較

表4のパラメータとAとBの比較を行なう。モデルは共に2層〔30×5, 50, 8〕で150msの時間範囲をニューラルネットへ入力する。音声資料はSM0について行なう。

#### (d) 単語音声への適用

単音節で学習したニューラルネットを用いて1桁数字音声15単語のセグメンテーションを行なう。モデルは2層〔30×5, 50, 8〕で150msの時間範囲をニューラルネットへ入力する。音声資料はSM0のパラメータAについてである。

### 3.3.4 実験結果

またニューラルネットの学習が収束しきっていないので、学習途中の傾向を示すにとどめる。

#### (e) セグメンテーション位置の細分化の効果

途中結果を見る限りにおいては細分化したモデル1の方がモデル2よりもセグメンテーション率が悪い。少なくとも当初予想したような細分化の効果は見られない。学習回数が違い、かつモデル1の方がニューラルネット規模が大きく収束しにくい点から、学習途中ではまだ結論付けられない。ニューラルネットへ一度に入力する時間が90msの場合をFig.13に示す。

#### (f) open実験

同一話者の発声時期の違うデータSM1, 別発声話者男女データSM2, SF1に対しては、語頭のようにあまり性能低下しない場合もあるが、全体的に同程度の割合でかなり悪くなっている。(Fig.14)

#### (g) 帯域分割の違いによる性能比較

全体的に殆ど変わらないが、CV境界の判定にBark scaleの帯域パワーの方が優位でありそうだ。

#### (h) 単語音声への適用

実験(f)以上に性能低下が激しく、単音節で学習したニューラルネットでは単

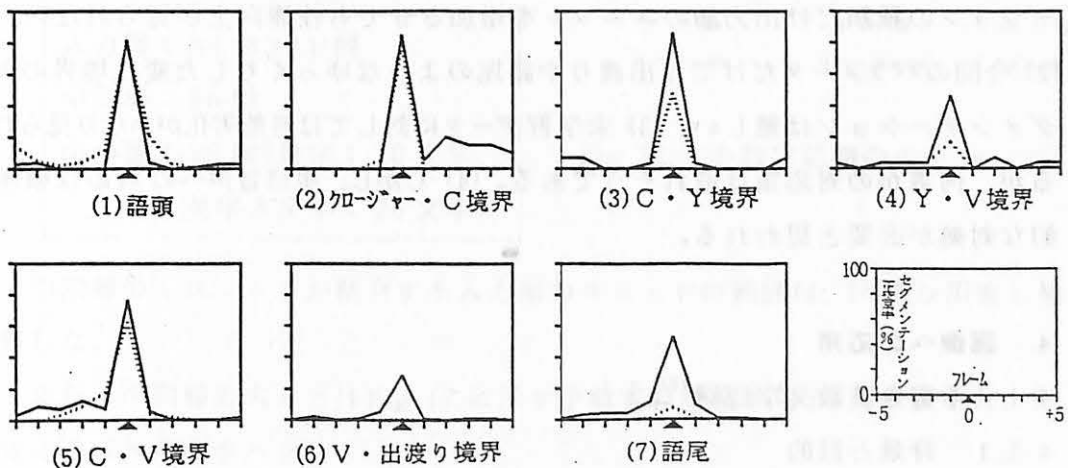


Fig.13 セグメンテーション位置の細分化の効果

音声資料：SM0のパラメータA

(.....) 細分化あり, 2層 [30×5, 50, 8]

(——) 細分化なし, 2層 [30×5, 50, 1]

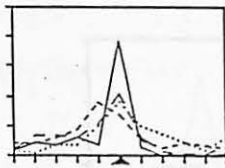


Fig. 14 単音節を用いたopen実験

音声資料：パラメータA

NNモデル2層 [30×5, 50, 1]

(—) SM0      (---) SM1  
(- - -) SM2      (.....) SF1

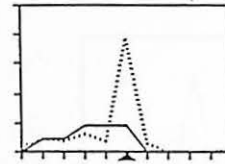


Fig. 15 単語音声を用いたopen実験

音声資料：パラメータA

NNモデル2層 [30×5, 50, 1]

(.....) SM0      (—) WM1

語・連続音声への適用が困難ではなかろうかと思わせる。(Fig.15)

(i) 全体的に

語頭・クロージャC境界・CY境界・CV境界はスペクトル変化が激しいため、容易にセグメンテーションできるが、YV境界・V出渡り境界・語尾ではFig.13に見られるように非常に不十分な結果である。今後の学習によって向上はするであろうが、ニューラルネットの代わりに視察で判断してもかなり難しいことから、本質的に今回用いた様なパワー情報だけでは限界があるものと思われる。

### 3.3.5 まとめ

学習途中の傾向からの判断ではあるが実験の傾向をまとめる。(1) セグメンテーションの種類だけ出力層のユニットを増加させても性能向上が見られない。(2) 今回のパラメータだけでは出渡りや語尾のようなゆっくりした変化境界のセグメンテーションは難しい。(3) 未学習データに対しては性能劣化がかなり見られるが、何等かの対応策は取れそうである。(4) しかし、単語音声への対応は根本的な対処が必要と思われる。

## 4. 画像への応用

### 4.1 手書き英数文字認識<sup>12)</sup>

#### 4.1.1 背景と目的

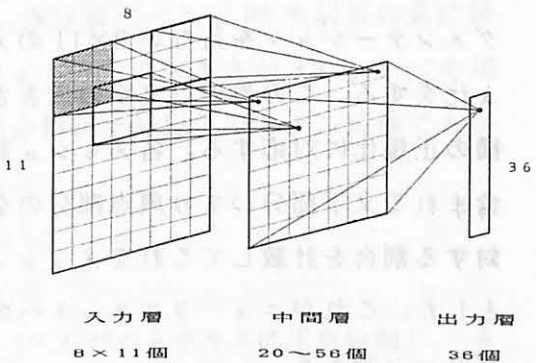
文字認識の分野でもニューラルネットを手書き数字<sup>13)</sup>、手書き漢字<sup>14)</sup>に適用し、従来手法に劣らない認識率が得られている。ここでは、以下の2点について認識性能を評価する。

第1は、計算量の削減と文字の部分空間情報の有効利用の可能性について検討である。文字は部分的な形状間には強い相関があると考えられる。漢字の場合の部首はその代表的なものである。この性質を用いればニューラルネットのモデル規模を小さくすることができる。3.1節のホルマント抽出で用いたモデル (Fig.7) を参考に、中間層と入力層の接続範囲を限定したニューラルネットを手書き英数字認識に適用し、モデルの規模と認識性能との関係を調べる。

第2は、学習人数についてである。不特定筆者対応に対するためには複数人数のデータを学習させ汎用性を持たせる必要がある。しかし多過ぎても性能は飽和状態になり学習効率が悪い。そこで、この学習人数と認識性能との関係を調べる。

#### 4.1.2 モデルの構成

今回認識実験に用いたモデルを、Fig.16に示す。中間層のユニットと入力層との接続については結合範囲を2次元的に変えることが可能な階層的ネットワークである。各層のユニット数は、以下のように設定した。



入力層：88 (8×11) 個
中間層：56 個
出力層：36 個 (数字：10 文字， 英字大文字：26 文字)

Fig.16 英数字認識のモデル

中間層の1ユニットが結合する入力層のユニットの範囲は、16通り用意し検討した。

また、中間層の大きさはFig.17に示す予備実験結果から56個に決めた。

#### 4.1.3 入力パターン

今回用いた手書き英数字データは、Fig.18に示すように、見本文字を参考に所定の枠内に数字10文字 (1~10)、大文字英字26文字 (A~Z) の合計36文字の80人分のデータを用いた。この文字をスキャナーで入力し64×63

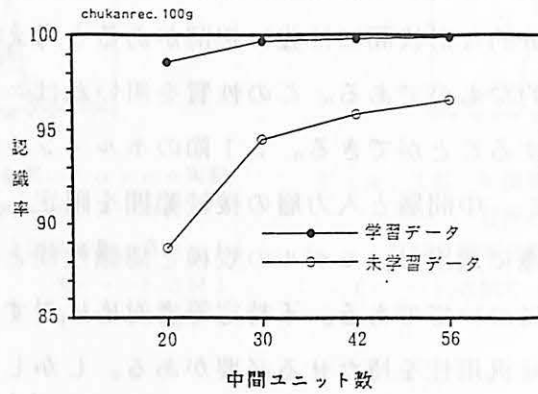


Fig.17 中間層の大きさと認識率

の二値マトリックスのパターンに変換したものを素データとする。次に文字部分のセグメンテーションを行ない8×11のメッシュ化をする。この処理は文字の大きさ・縦横の正規化に対応する。各メッシュ領域に含まれる文字部分つまり黒点部分の全体に対する割合を計数してこれをメッシュ特徴とした。これがニューラルネットへの入力パターンである。



Fig.18 手書き文字の例

#### 4.1.4 実験I：ウィンドウの大きさと認識率

##### (a) 実験方法

40人を学習データとし残りの40人を評価用データとする。ウィンドウの大きさは(3×3)～(8×11)の16通りに行ない認識率との関係を調べる。

##### (b) 実験結果

Fig.19-1, Fig.19-2に100回学習・200回学習を行なった時の認識率及び結合数を示す。結合数は、ウィンドウの大きさが8×11の場合(中間層の各ユニットが全入力ユニットと結合した場合)を1としてそれとの比(結合数比)で表している。

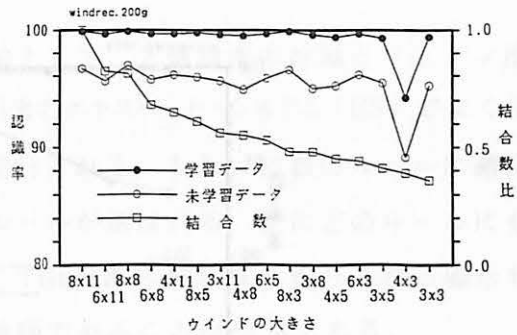
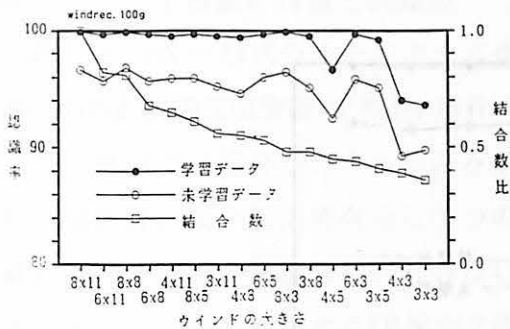


Fig.19-1 ウィンドウの大きさ  
と認識率（100回学習時）

Fig.19-2 ウィンドウの大きさ  
と認識率（200回学習時）

学習データ，未学習データともに，200回学習のネットワークに比べると100回学習の方は，ウィンドウの大きさによる認識率のばらつきが目立つものの，どちらの場合もウィンドウの大きさのある程度小さくしても，それほど認識率の劣化はみらず，学習データで99%以上，未学習データで96%前後の高認識率を得た。特に，未学習データに対するウィンドウの大きさが8×3，6×3の場合などウィンドウの取り方によっては，結合数が半分以下のモデル規模でもかなり高い認識率を示している。

#### 4.1.5 実験Ⅱ：学習人数と認識率

##### (a) 実験方法

open実験では，80人分の学習データを10人ずつ8クラスに任意分割し，そのうち1クラスを評価用データ，他の7クラスの幾つかを学習データとする。データの片寄りを除くため，評価用クラスと学習用クラスを4通りに替えて評価し，4回の平均値を求めた<sup>15)</sup>。close実験でも同様のデータ分割を行ない，学習データを評価用にも用いた。ウィンドウの大きさは8×11，学習回数は100回である。

##### (b) 実験結果

Fig.20に評価実験結果を示す。open実験では，学習データ人数の増加に伴い認識率が向上するが40人程度を境に飽和している。一方，close実験は学習人数に関係無く100%に近い認識性能が続いている。

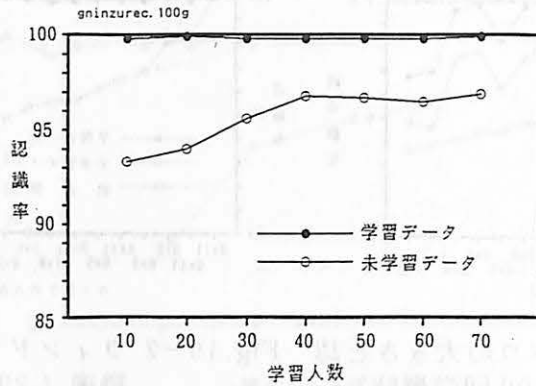


Fig.20 学習データ人数と認識率

#### 4.1.6 考察

認識性能を劣化させることなくウィンドウの大きさを小さくできたことから、このニューラルネットが中間層と入力層の結合範囲を限定したことによって文字の部分空間情報を有効に利用していることを示唆していると考えられる。

また、学習データ人数の実験については、40人程度で性能が飽和状態になるという実験結果を得たが、認識対象文字が36英数字であることと今回のモデル規模とに依存すると思うのが妥当である。いずれにせよ、不特定筆者用の文字認識システムを設計する上での設計指針が与えられたことには違いない。

#### 4.1.7 まとめ

中間層と入力層の接続範囲を限定したニューラルネットを手書き英数字認識に適用し、演算量の削減と不特定筆者用の文字認識システムにおける必要学習データ人数について検討した。その結果、中間層と入力層の接続範囲を限定し結合数を削減してもほとんど認識率の劣化はなく学習データで99%以上、未学習データで96%前後の高認識率を得た。また、英数字36文字認識の実験では、40人程度で認識性能が飽和した。

### 5. 制御への応用

#### 5.1 ニューラルネット駆動型ファジィ推論<sup>16)~20)</sup>

##### 5.1.1 背景と目的

(a) ファジィ推論の特徴と問題点

IF…THEN…形式のルールベース推論として、2値論理の推論とファジィ推論(多値論理/連続値論理)がある。前者は旧来のエキスパート・システム(ES)でよく使われた推論で、適合したルールのみが実行される。もし、複数のルールに適合した場合は、競合解消で有力な1つのルールが選ばれる。逆にどのルールにも適合しない場合、推論は行なわれない。THEN部に<sup>確</sup>信度を取り入れる場合もあるが、ルールを選択するIF部が2値論理であることが特徴である。

これに対し、確定的でない入力変数の推論ルールを持つものがファジィ推論である。ルールを選択するIF部に曖昧さを包含しているため人間の思考を直接推論ルールとすることができ、推論ルールの設計が容易である。また、IF部の曖昧さのため、複数ルールに適合することになる。各ルールのメンバーシップ関数から帰属度を求め、全てのTHEN部を実行して帰属度との重み付けによって最終推論結果を求める。曖昧さを扱うことによって、推論対象を拡大し、推論ルールを柔軟にすることができる。最近ではファジィ推論ルールベースのESも増えている。

しかし、設計が容易であることと、実現が容易であることは必ずしも同一ではない。例えば、「IF室温がほぼ標準温度 and 室温が緩やかに上昇 and 人の不快感はあまりない, THEN 火力を中位に落とす」とファジィ推論ルールを設計することは容易であるが、実際のセンサーからの入力値やTHEN部の制御量は1個の確定値であり、これらの橋渡しするメンバーシップ関数の決定が問題となる。具体的には、従来のメンバーシップ関数の設計ではIF部の変数(前述の例では温度と湿度)が独立であると仮定し、Fig.22のようなファジィ推論ルール分割を行なって問題を単純化し、試行錯誤で、メンバーシップ関数の形状を決定していた。しかし、現実には制御における「温度」と「湿度」は相互に相関関係があると考えた方が妥当であることからわかるように、Fig.21のようなファジィ推論ルール分割を考えるべきで、Fig.21の多次元空間上の「歪んだメンバーシップ関数を設計すべきなのである。ところが、この変数が3個4個と増えるとそのような多次元空間上で経験的にメンバーシップ関数を決定することなどできなくなってしまう。これが従来のファジィ推論の第1の課題であ



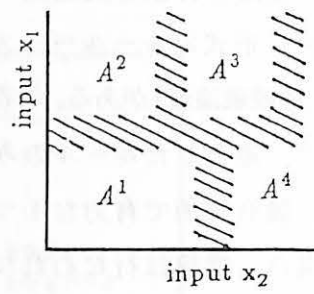
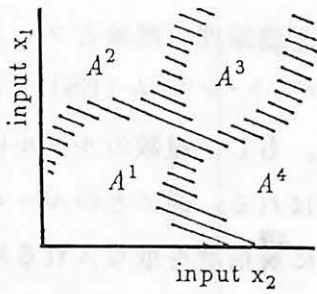


Fig.21 提案方法のルールのファジィ分割      Fig.22 従来のルールのファジィ分割

る。

第2の課題は、推論ルールを環境変化に伴って修正する必要が生じた場合、自動的に自己修正するような学習機能を持たないことである。

(b) 目的

ニューラルネットには学習機能・非線形性の特長があり、ルールがよく分からない複雑なシステムに対しては入出力関係だけで学習によって最適な分離が可能である。これはESやファジィ推論にはない大きな特長である。反面、内部形成されたルールが分からないため、ルールベースのような不変性というか、納得できる論理性はない。学習データが一般性を持った分布をしていけばよいが、偏った学習データには「きれいに」偏ってしまう。この点がルールベースとニューラルネットとの住み分けができる点である。

ニューラルネット駆動型ファジィ推論は、推論部分は論理性のあるファジィ推論を行ない、試行錯誤で行なっていたメンバーシップ関数の決定をニューラルネットで行なうシステムである。

5.1.2 ファジィ推論の定式化 (ファジィモデリングを例に)

システムの制御等ではTHEN部が確定的な制御操作量を推定することになる。このような手法にファジィモデリング<sup>21), 22)</sup>があり、「IF  $\mathbf{X} \in \mathbf{A}^s$ , THEN  $y = F(\mathbf{x})$ 」の形式をしている。ここで、 $\mathbf{X}$ は入力変数のベクトル、 $\mathbf{A}^s$ はFig.21の分割されたルール空間、 $F(\cdot)$ は制御操作量推定関数である。本提案方法では、 $\mathbf{A}^s$ がニューラルネットによってFig.21のようにファジィ分割され、 $F(\cdot)$ がニューラルネットそのものになる。以下その手順を示す。

【手順1】観測値  $y_i$  を出力，各入力数  $x_j, j = 1, 2, \dots, k$  を入力とするニューラルネットを用いて誤差二乗和を評価指標とする変数減少法により，出力層に関連がある入力変数  $x_j, j = 1, 2, \dots, m, m \leq k$  のみを選択する。これは外乱となる入力変数を取り除き，従来の相関関係の大きい入力変数を選択することに相当する。

【手順2】入力データ  $(x_i, y_i)$  をモデル推定の学習用データ（以下，TRDと記す。  $n_t$  個）と推定モデルの評価用データ（以下，CHDと記す。  $n_c$  個）とに分割する。（  $n = n_c + n_t$  ）

【手順3】TRDを通常のクラスタリング手法を用いて最適な  $r$  分割を行う。 $r$  分割された学習データの各組を  $R^s, s = 1, 2, \dots, r, R^s$  の学習データを  $(x_i^s, y_i^s), i = 1, 2, \dots, n_t$ ，  $r$  分割は推論ルール数を  $r$  個にすることになる。

【手順4】IF部構造の同定を行う。 $x_i$  を入力層の入力値に割り付け，出力層の出力値として，

$$w_i^s = \begin{cases} 1 & : (x_i, y_i) \in R^s \\ 0 & : (x_i, y_i) \notin R^s \end{cases}$$

for  $i = 1, \dots, n_t, s = 1, \dots, r$

を割り付ける。学習によって  $w_i^s$  を推定するニューラルネットを固定し，各学習データ  $(x_i, y_i)$  の各  $R^s$  に属する度合い  $w_i^s$  を推定する。IF部のメンバシップ関数は推定値  $\hat{w}_i^s$  で定義する。即ち，

$$\mu_A^s(x_i, y_i) = \hat{w}_i^s, \quad i = 1, 2, \dots, n_t$$

とする。 $\mu_A^s(x_i, y_i) > 0$  となる学習データを新たに，  $(x_i^s, y_i^s), i = 1, 2, \dots, (n_t)^s$  とする。

【手順5】THEN部構造の同定を行う。各推論ルールのTHEN部構造をニューラルネットの入出力関係で表わし，TRDの入力値  $x_{i1}^s, \dots, x_{im}^s, i = 1, 2, \dots, (n_t)^s$  と出力値  $y_i^s$  を割り付ける。学習によって制御操作量を推定するニューラルネットを同定する。次に，得られたニューラルネットにCHDの入力値  $x_{i1}^s, \dots, x_{im}^s, i = 1, 2, \dots, (n_t)^s$  を代入し，誤差二乗和  $\theta_m^s$  を求める。

$$\theta_m^S = \sum_{i=1}^{(n_c)^S} (y_i^S - \hat{y}_i^S)^2 \quad (4)$$

【手順6】変数減少法を用いる。各推論ルールのTHEN部の制御操作量を推定するニューラルネットのm個の入力変数の中で、任意の1個の入力変数を取り除き、手順5と同様にTRDを用いてTHEN部毎にニューラルネットを同定する。次に、CHDを用いた場合の制御操作量の推定誤差二乗和  $\theta_{m-1}^{SP}$  を計算する。

$$\theta_{m-1}^{SP} = \sum_{i=1}^{(n_c)^S} (y_i^S - \hat{y}_i^S)^2, \quad p = 1, 2, \dots, m \quad (5)$$

(1)式と(2)式を比較し、

$$\theta_m^S > \theta_{m-1}^{SP} \quad (6)$$

となる場合は取り除いた入力変数  $x^P$  の重要度は低いと考えられるので、 $x^P$  を捨てる。

【手順7】残りの入力変数をm個として手順5と同様の操作を行う。以下、手順5～手順6を繰り返し、(6)式が全ての入力変数に対して成立しなくなった場合に計算を停止する。 $\theta^S$ が最小モデルとなるモデルが最適なニューラルネットである。

手順1～手順7により、各推論ルール毎のIF部とTHEN部が決定され、ファジィモデルの構造同定が終了する。

【手順8】制御操作量  $y_i^*$  は次式により得られる。

$$y_i^* = \frac{\sum_{s=1}^r w_i^S \tilde{y}_i^S}{\sum_{s=1}^r w_i^S}, \quad i = 1, 2, \dots, n_c \quad (7)$$

ただし、 $\tilde{y}_i^S$  は手順7で得られた最適なニューラルネットにCHDを代入した推定値を示す。

上記のうち非線形メンバーシップ関数を自動的に決定する手順についてさらに詳しく説明しよう。

似通った入力データには同一制御ルールが適用されると考えるのは自然である。そこで、手順3で学習データをクラスタリングする (Fig.23(a)) これらのク

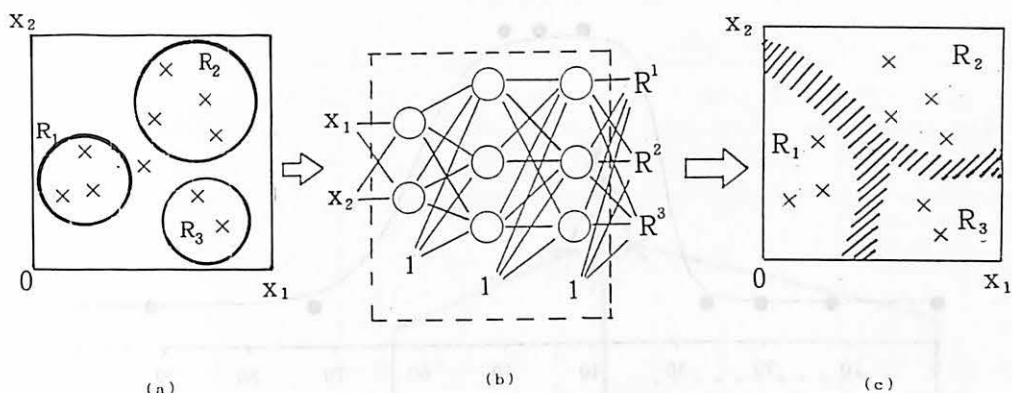


Fig.23 ニュラルネットによるファジィ推論ルールの分割

ラスが1つ1つのルールに対応している。例えば、「IF  $x_1$  が小さく、かつ、 $x_2$  が大きければ、THEN ...」といったルールである。Fig.23 (a)の場合このルールが  $R_1 \sim R_3$  の3個ある。これらのクラスの代表的な点（入力データの組）では各ルールにはほぼ100%従うが、境界に近づくにつれ、複数ルールに少しずつ適合していくことになる。この帰属度合いがメンバーシップ関数であり、その形状は例えばFig.23 (c)である。これは上から見たメンバーシップ関数の図で斜線部はメンバーシップ関数が交差している領域である。

さてこのようなメンバーシップ関数を作る手順が手順4である。即ち、Fig. 23 (b)のニューラルネットを用意し、入力にIF部の変数値（図の場合  $x_1$  と  $x_2$ ）を割り当て、出力にその入力値が属するルール番号に1、その他に0を割り当てて学習させる。ニューラルネットの重要な特長の1つに、類似入力には類似出力が対応する点がある。学習にはFig.23 (a)の×点のみが使われたが、その近傍であれば同じ出力、つまり同じ制御ルールに従う旨出力してくる。更に、学習を終えたニューラルネットは各ルールの狭間の入力データに対しては、それぞれのルールへの帰属割合の最適バランス値を出力することになる。これがメンバーシップ値で、言い換えれば、このニューラルネットは各ルールの全てのメンバーシップ関数を内部に包含したといえる。

ところで、実際ニューラルネットでメンバーシップ関数が作れるのかどうか疑問もあろうかと思う。その回答がFig.24である。「明らかに中年でない年齢」

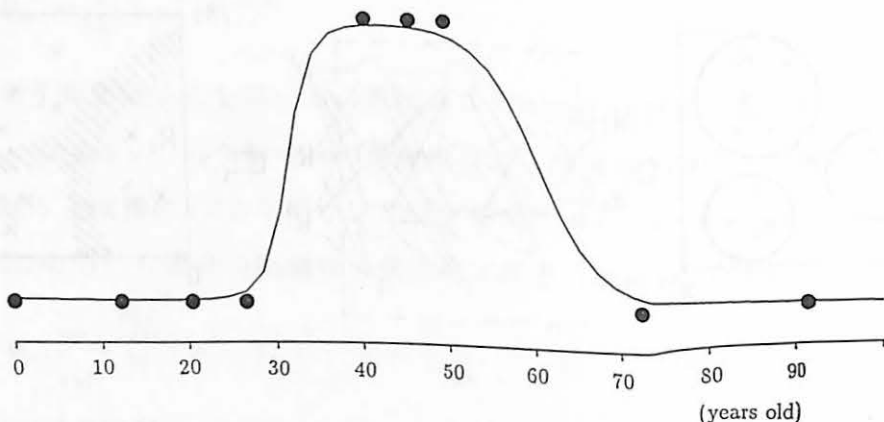


Fig.24 ニューラルネットによる「中年」のメンバーシップ関数の決定

と「明らかに中年の年齢」との9データでニューラルネットを学習させた結果、図の曲線のようなメンバーシップ関数が内部に形成された。この場合、変数は年齢という1次元なので経験と勘でも容易に図のようなメンバーシップ関数が設計できるが、これがFig.21のように2次元になったり多次元空間上の超曲面となると、もう従来のような経験では設計できなくなる。

### 5.1.3 システム構成

以上の定式化でニューラルネット駆動型ファジィ推論が実行できることになった。この推論を実行するシステム例をFig.25に示す。図中左端のニューラルネット  $NN_{mem}$  は手順4で作成したIF部に相当するメンバーシップ決定ニューラルネット (Fig.23(a)) で、 $NN_1 \sim NN_r$  は手順6～手順8で作成したTHEN部のニューラルネットである。THEN部の出力はIF部の帰属度を重み付けられ最終出力値を求める (手順9)。

この基本システムは展開可能で、例えば、入力を1つの確定値ではなくファジィ数であるならば、ニューラルネットの複数の入力を1つのファジィ数入力に対応させればよい。また、出力値がファジィ数であるならば、その逆にすればよい。

### 5.1.4 具体的適用例

定式化したニューラルネット駆動型ファジィ推論の具体的適用を示すために、菅野が提案したファジィモデリングで用いた簡単な数値例<sup>21), 22)</sup>を用いて、上述

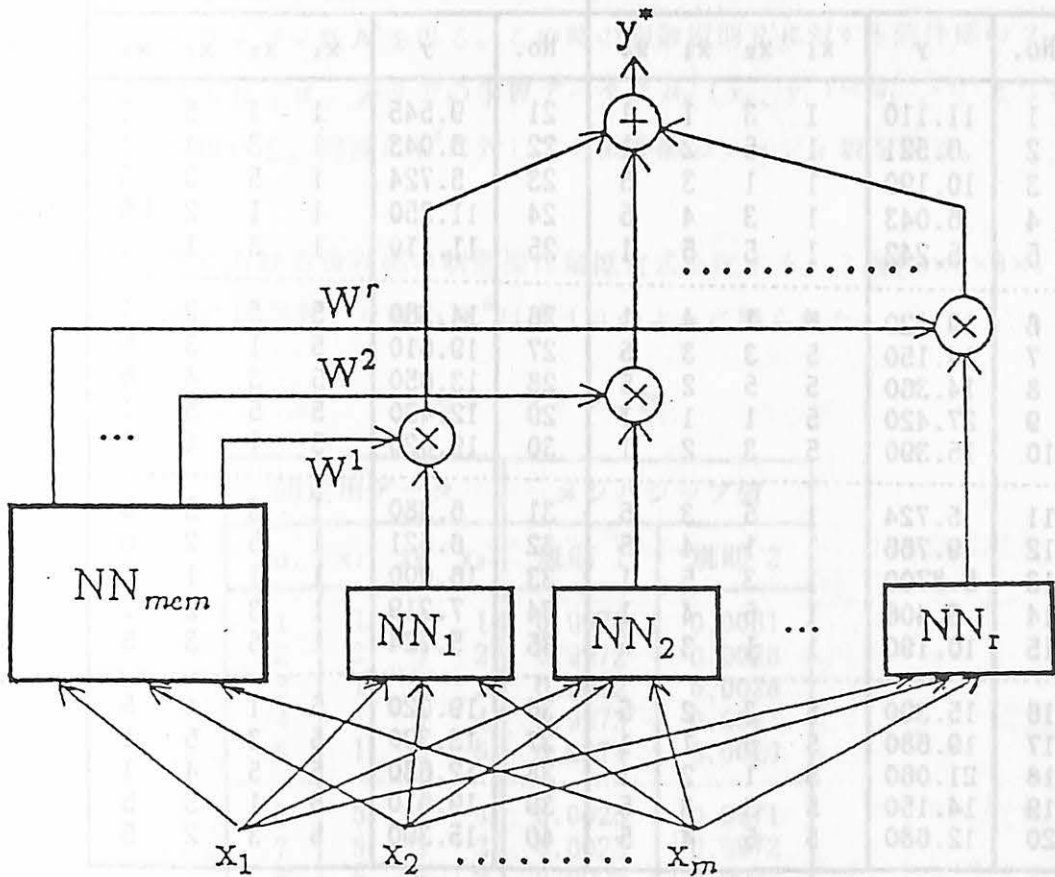


Fig.25 ニューラルネット駆動型ファジィ推論システム

手順に従ってファジィ推論を行う。

【手順 1, 2】

表 6 に入力データを示す。データ番号 1-20 は学習用 (TRD) であり、データ番号 21-40 は評価用 (CHD) である。したがって  $n_t = n_c = 20, m = 2$  となる。3 層  $[3 \times 3 \times 3 \times 1]$  モデルで 15000 回 学習して入力変数の選択を行なった結果、表 7 の結果が得られた。

これにより、入力変数  $x_4$  を省いても推定にはほとんど影響がないと判断し、

表6 入出力データ

同定用データ					評価用データ						
No.	y	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	No.	y	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>
1	11.110	1	3	1	1	21	9.545	1	1	5	1
2	6.521	1	5	2	1	22	6.043	1	3	4	1
3	10.190	1	1	3	5	23	5.724	1	5	3	5
4	6.043	1	3	4	5	24	11.250	1	1	2	5
5	5.242	1	5	5	1	25	11.110	1	3	1	1
6	19.020	5	1	4	1	26	14.360	5	5	2	1
7	14.150	5	3	3	5	27	19.610	5	1	3	5
8	14.360	5	5	2	5	28	13.650	5	3	4	5
9	27.420	5	1	1	1	29	12.430	5	5	5	1
10	15.390	5	3	2	1	30	19.020	5	1	4	1
11	5.724	1	5	3	5	31	6.380	1	3	3	5
12	9.766	1	1	4	5	32	6.521	1	5	2	5
13	5.8700	1	3	5	1	33	16.000	1	1	1	1
14	5.406	1	5	4	1	34	7.219	1	3	2	1
15	10.190	1	1	3	5	35	5.724	1	5	3	5
16	15.390	5	3	2	5	36	19.020	5	1	4	5
17	19.680	5	5	1	1	37	13.390	5	3	5	1
18	21.060	5	1	2	1	38	12.680	5	5	4	1
19	14.150	5	3	3	5	39	19.610	5	1	3	5
20	12.680	5	5	4	5	40	15.390	5	3	2	5

以後の実験では用いない。

【手順3】

TRDを通常のクラスタリング手法を用いて分割する。分割された各学習データは次のようになる。

表8 学習データのルール分割

制御規則	学習データ番号
R <sup>1</sup>	1, 2, 3, 4, 5, 11, 12, 13, 14, 15
R <sup>2</sup>	6, 7, 8, 9, 10, 16, 17, 18, 19, 20

表7 ニューラルネットによる変数減少法の結果

	誤差二乗和
全変数使用した場合	0.0007
x <sub>1</sub> を取り除いた場合	0.2936
x <sub>2</sub> を取り除いた場合	0.1482
x <sub>3</sub> を取り除いた場合	0.0872
x <sub>4</sub> を取り除いた場合	0.0019

【手順4】

学習データ  $(x_i, y_i), i = 1, 2, \dots, 20$  の  $A^S$  に属する度合い  $\hat{w}_i^S = \{0, 1\}$  を  $\hat{w}_i^S \in [0, 1]$  の値で推定するために3層  $[3 \times 3 \times 3 \times 2]$  モデルを5000回学習し、前件部のファジィ数  $A^S$  を得る。この時の制御規則  $R^1$  に対する前件部のファジィ数  $A^1$  を表9に示す。表9での学習データは  $\mu_R^1(x_i, y_i) = w_i^1 > 0$  となる入力データを用いる。同様に  $R^2$  に対しても前件部のファジィ数を得る。

【手順5】

各制御規則における後件部の制御操作量推定式を定する。3層  $[3 \times 8 \times 8 \times 1]$  モデルを20000回学習した後の  $\theta_4^S$  は表10のように得られた。

表9 ファジィ数

同定用データ				メンバシップ値	
No.	$x_1$	$x_2$	$x_3$	規則 1	規則 2
1	1	3	1	0.9970	0.0031
2	1	5	2	0.9972	0.0028
3	1	1	3	0.9972	0.0028
4	1	3	4	0.9973	0.0027
5	1	5	5	0.9974	0.0026
-----					
6	5	1	4	0.0028	0.9971
7	5	3	3	0.0028	0.9972
8	5	5	2	0.0027	0.9972
9	5	1	1	0.0027	0.9973
10	5	3	2	0.0027	0.9973
-----					
11	1	5	3	0.9973	0.0028
12	1	1	4	0.9973	0.0027
13	1	3	5	0.9974	0.0026
14	1	5	4	0.9973	0.0027
15	1	1	3	0.9972	0.0028
-----					
16	5	3	2	0.0027	0.9973
17	5	5	1	0.0027	0.9973
18	5	1	2	0.0027	0.9973
19	5	3	3	0.0028	0.9972
20	5	5	4	0.0029	0.9971



【手順 6, 7】

制御規則  $R^S$  の後件部の構造モデルから任意の 1 個の入力変数を取り除いた場合の  $\theta_3^S$  を求める。3 層 [  $2 \times 8 \times 8 \times 1$  ] モデルを 10000 ~ 20000 回学習した結果、制御規則  $R^1, R^2$  に対して表 11 の誤差二乗和が得られた。

表 10 出力誤差

制御規則 1 : $\theta_4^1$	27.86
制御規則 2 : $\theta_4^2$	1.93

表 11 変数減少後の出力誤差

	規則 1 : $\theta_3^1$	規則 2 : $\theta_3^2$
$x_1$ を取り除いた場合	42.84	0.93
$x_2$ を取り除いた場合	74.71	119.61
$x_3$ を取り除いた場合	55.27	73.28

各制御規則毎に手順 5 と手順 6 を比較すると

$$\text{全ての } \theta_3^1 > \theta_4^1 (= 27.86)$$

$$x_1 \text{ を取り除いた場合 } \theta_3^2 < \theta_4^2 (= 1.93)$$

となっているので、制御規則 1 に対しては手順 5 のニューラルネットを後件部モデルとする。制御規則 2 に対しては更に計算を続け、第 2 段階目の繰り返し計算で終了する。(  $x_2, x_3$  ) 入力のニューラルネットを後件部モデルとする。得られたファジィモデルは

$$R^1 : \text{IF } \mathbf{X} = (x_1, x_2, x_3) \text{ is } \mathbf{A}^1 \text{ THEN } y^1 = \text{NN}_1(x_1, x_2, x_3)$$

$$R^2 : \text{IF } \mathbf{X} = (x_1, x_2, x_3) \text{ is } \mathbf{A}^2 \text{ THEN } y^2 = \text{NN}_2(x_2, x_3)$$

となる。得られた  $R^1$  の規則クラスリングを  $x_2 - x_3$  平面図として Fig. 26 に示す。また、(7) 式の  $y_i^*$  を表 12 に示す。

5. 1. 5 応用 例

ニューラルネット駆動型ファジィ推論の有用性を示すために、3 つの事例に適用する。第 1 は 5. 1. 4 節で用いた数式モデルの推定、第 2 は大阪湾での COD 濃度の推定、第 3 はカップ形ダイヤモンド砥石によるセラミックスの平面研削で

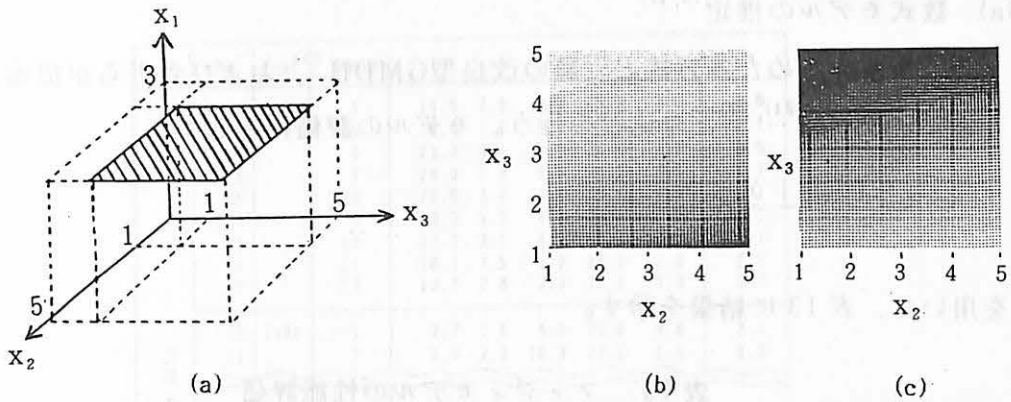


Fig.26 制御規則のメンバシップ関数の超平面

表 12 制御操作量の推定結果

同定用データ					評価用データ				
No.	y	y*	w <sub>1</sub> <sup>1</sup>	w <sub>1</sub> <sup>2</sup>	No.	y	y*	w <sub>1</sub> <sup>1</sup>	w <sub>1</sub> <sup>2</sup>
1	11.110	11.136	0.9970	0.0031	21	9.545	8.882	0.9974	0.0027
2	6.521	6.534	0.9972	0.0028	22	6.043	6.140	0.9973	0.0027
3	10.190	10.210	0.9972	0.0028	23	5.724	5.712	0.9973	0.0028
4	6.043	6.140	0.9973	0.0027	24	11.250	10.547	0.9971	0.0030
5	5.242	5.370	0.9974	0.0026	25	11.110	11.136	0.9970	0.0031
6	19.020	18.995	0.0028	0.9971	26	14.360	14.334	0.0027	0.9972
7	14.150	14.134	0.0028	0.9972	27	19.610	19.061	0.0028	0.9972
8	14.360	14.334	0.0027	0.9972	28	13.650	13.918	0.0029	0.9971
9	27.420	27.373	0.0027	0.9973	29	12.430	12.293	0.0030	0.9969
10	15.390	15.383	0.0027	0.9973	30	19.020	18.995	0.0028	0.9971
11	5.724	5.712	0.9973	0.0028	31	6.380	7.178	0.9972	0.0028
12	9.766	9.791	0.9973	0.0027	32	6.521	6.534	0.9972	0.0028
13	5.8700	5.747	0.9974	0.0026	33	16.000	11.239	0.9969	0.0032
14	5.406	5.450	0.9973	0.0027	34	7.219	9.018	0.9971	0.0029
15	10.190	10.210	0.9972	0.0028	35	5.724	5.712	0.9973	0.0028
16	15.390	15.383	0.0027	0.9973	36	19.020	18.995	0.0028	0.9971
17	19.680	19.652	0.0027	0.9973	37	13.390	13.892	0.0030	0.9970
18	21.060	21.046	0.0027	0.9973	38	12.680	12.672	0.0029	0.9971
19	14.150	14.134	0.0028	0.9972	39	19.610	19.061	0.0028	0.9972
20	12.680	12.672	0.0029	0.9971	40	15.390	15.383	0.0027	0.9973

の仕上面粗さの予測である。

(a) 数式モデルの推定<sup>16, 17)</sup>

5. 1. 4 節で求めた本方式と近藤の改良型GMDH<sup>23)</sup>、および菅野らが提案したファジィモデル<sup>21), 22)</sup>との比較を行う。モデルの評価指として、

$$D = \frac{1}{n} \sum \frac{|y_i - y_i^*|}{y_i}$$

を用いて、表 13 に結果を示す。

表 13 ファジィモデルの性能評価

	同定用データ	評価用データ
近藤 GMDH	4.7	5.7
菅野モデル 1	1.5	2.1
菅野モデル 2	1.1	3.6
本提案モデル	0.47	4.79

(b) 大阪湾でのCOD濃度の推定<sup>17), 20)</sup>

大阪湾でのCOD濃度を推定する。表 14 に示す入力データとして 1976 年 4 月から 1979 年 3 月までの測定データが得られている。入出力変数は次の通りである。

$y$  : COD 濃度 (ppm)

$x_1$  : 水温 (°C)

$x_2$  : 透明度 (m)

$x_3$  : DO 濃度 (ppm)

$x_4$  : 塩分濃度 (%)

$x_5$  : 濾過したCOD濃度 (ppm)

藤田らはまず、濾過したCOD濃度  $x_5$  を拡散シミュレーションの結果から推定し、この濾過したCOD濃度をGMDHモデルの入力変数に用いて、出力のCOD濃度  $y$  を推定している<sup>24)</sup>。

本論文では同じデータを用いて、ニューラルネット駆動型ファジィ推論により、出力のCOD濃度  $y$  を推定する。

表 14 入出力データ

	No.	Year	Month	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	y	
Training Data	1	1976	4	14.9	2.0	9.3	29.8	3.0	3.8	
	2		5	16.6	1.7	5.8	23.1	3.3	3.7	
	3		6	21.3	2.1	9.1	24.0	2.2	3.0	
	4		7	24.3	2.9	7.0	24.1	3.4	4.7	
	5		8	26.6	1.7	4.8	21.8	2.9	4.1	
	6		9	23.2	1.3	4.7	28.7	1.6	3.1	
	7		10	22.2	2.5	4.5	25.6	1.3	3.1	
	8		11	18.1	2.5	5.9	29.0	2.0	2.5	
	9		12	13.7	2.8	7.9	25.9	3.3	3.3	
	10	1977	1	7.7	2.6	9.3	29.0	0.8	2.3	
	11		2	6.9	2.3	10.3	27.5	1.5	3.3	
	21			12	15.2	2.5	6.5	28.4	1.2	2.9
	22	1978	1	10.0	3.0	8.6	28.3	1.1	3.1	
	23		2	8.7	1.9	8.6	27.8	1.8	3.7	
	24		3	7.5	1.5	9.8	27.8	1.6	2.9	
	25		4	12.5	1.7	10.0	25.9	2.2	4.7	
	26		5	17.7	1.1	11.0	27.8	2.5	5.4	
	27		6	21.4	1.9	8.6	23.0	2.8	5.5	
	28		7	26.6	2.3	4.8	22.2	2.5	3.7	
	29		8	28.7	1.1	8.2	28.7	3.8	8.1	
	30		9	26.7	2.5	8.6	31.3	2.0	4.1	
	31		10	23.2	2.5	5.4	30.9	1.6	2.9	
	32		11	19.1	3.0	5.1	30.5	1.7	3.1	
	33		12	14.2	3.0	7.1	29.1	2.1	2.3	
	Checking Data	1	1979	1	12.5	2.5	7.4	29.3	1.8	2.9
		2		2	9.3	2.4	9.4	28.9	2.2	2.8
		3		3	9.5	2.5	9.4	30.1	2.1	2.7
		4		4	12.9	2.5	9.1	29.2	2.1	4.3
		5		5	16.0	2.1	10.0	25.8	2.9	4.9
		6		6	19.7	1.6	10.0	22.9	3.0	5.9
		7		7	22.8	3.0	5.0	22.2	2.0	3.2
		8		8	28.2	1.0	10.0	25.9	3.3	8.4
		9		9	24.6	2.8	6.4	30.7	2.1	4.3
10		10		22.2	2.3	5.9	30.1	2.6	3.6	
11		11		17.9	2.5	6.7	27.0	2.0	3.0	
12		12		14.3	4.0	7.5	28.1	1.5	3.1	

推定のためのニューラルネット駆動型ファジィ推論モデルを得るために次の条件で実験を行なった。

1) 同定用データとして、1976年4月～1978年12月までの32個を用い、評価用データとして、1979年1月～12月までの12個を用いる。

2) ニューラルネットワークの構造を

入力変数の選択用として、3層  $[4 \times 12 \times 12 \times 1]$ 、

前件部の構造決定用として、3層  $[5 \times 12 \times 12 \times 2]$ 、

後件部の構造決定用として、3層  $[5 \times 12 \times 12 \times 1]$ 、

とする。

3) 学習回数を 1500 ~ 2000 回とする。

結果として得られるニューラルネット駆動型ファジィ推論は次のようになる。

$R^1$ ; IF  $\mathbf{X} = (x_1, \dots, x_5)$  is  $\mathbf{A}^1$   
 THEN  $y^1 = NN_1(x_1, x_2, x_3, x_4, x_5)$

$R^2$ ; IF  $\mathbf{X} = (x_1, \dots, x_5)$  is  $\mathbf{A}^2$   
 THEN  $y^2 = NN_2(x_1, x_2, x_3, x_4, x_5)$

結果の推定値  $y_1^*$  , および観測データを Fig.27 に示す。

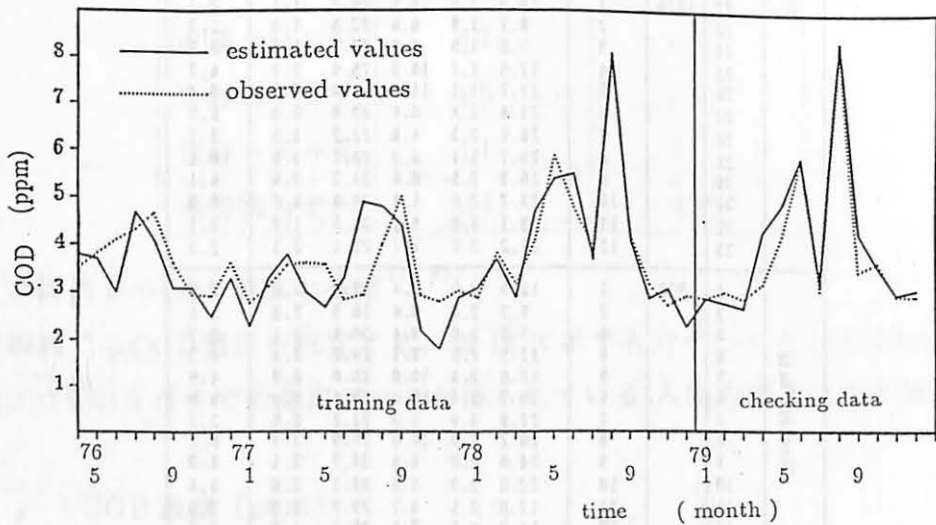


Fig.27 大阪湾のCOD濃度

次に、GMGHによるCOD濃度の推定結果と本論文で提案したファジィ推論を用いた推定結果との比較を行なう。モデルの標価指標として、

$$D = \sum_i (y_i - y_i^*)^2$$

を用いて、表 15 に結果を示す。

表 15 評価結果

	学習データ	未学習データ
<del>GMGH</del>	3.63	2.04
提案方式	3.52	1.58

(c) セラミックスの平面研削の仕上面粗さの予測

次に、ニューラルネット駆動型ファジィ推論をセラミックスの平面研削における仕上面粗さの予測に適用した例を示す。入力変数は下記の通りである。

$y$  : 仕上面粗さ ( $\mu\text{m}$ )

$x_1$  : 砥石周速 ( $\text{m}/\text{min}$ )

$x_2$  : 送り ( $\text{mm}/\text{min}$ )

$x_3$  : 切込み ( $\text{mm}$ )

$x_4$  : 粒度

$x_5$  : 集中度

上記の入力変数を用いて、様々な研削条件を設定し、マシニング・センターを使用してセラミックスの平面研削を行う。研削後は表面形状測定装置を用いて、被研削材の仕上面粗さを測定する。本論文ではニューラルネット駆動型ファジィ推論により、入力変数から出力の仕上面粗さ  $y$  を推定する。

推定モデルを得るために次の条件で実験を行なった。

1) 同定用データとして、次式で示される分散値の大きい 13 個を用いる。

$$S^2 = \Sigma ((x_{ij} - \bar{x}_j) / \sigma_j)^2$$

$$\sigma_j^2 = (\Sigma (x_{ij} - \bar{x}_j)^2) / n$$

2) 評価用データとして残りの 8 個を用いる。

3) ニューラルネットワークの構造、および学習回数は(b)と同じとする。

結果として得られるニューラルネット駆動型ファジィ推論は次のようになる。

$R^1$  ; IF  $\mathbf{X} = (x_1, \dots, x_5)$  is  $\mathbf{A}^1$

THEN  $y^1 = \text{NN}_1(x_1, x_2, x_4, x_5)$

$R^2$  ; IF  $\mathbf{X} = (x_1, \dots, x_5)$  is  $\mathbf{A}^2$

THEN  $y^2 = \text{NN}_2(x_1, x_2, x_3, x_4, x_5)$

$R^3$  ; IF  $\mathbf{X} = (x_1, \dots, x_5)$  is  $\mathbf{A}^3$

THEN  $y^3 = \text{NN}_3(x_2, x_4)$

結果の推定値  $y_i^*$  , および観測データを Fig.28 に示す。

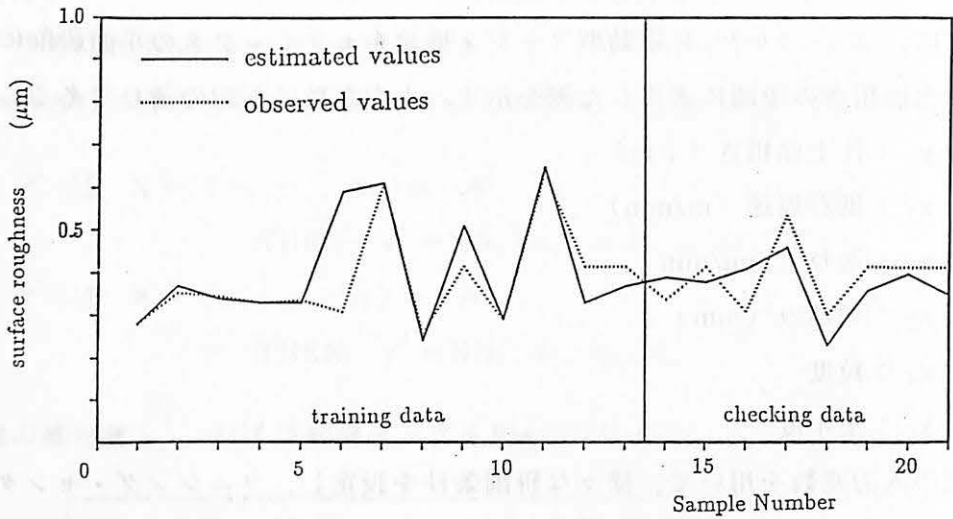


Fig. 27 セラミックス表面加工の粗さ推定

#### 5. 1. 6 考察とまとめ

3つの応用例を通じて、ニューラルネットの特徴がよく現れている。即ち、学習データには大変うまく適用しやすい点である。これは、未学習データに対しては欠点にもなる。学習データが偏りのない一般性のあるデータならばよいが、偏ったデータには「うまく」偏ってしまうためである。これを防ぐためには多くの学習データを用い統計的に偏りがないようにすることが望ましい。3つの応用例を見ると、全般にデータ数が不足ぎみであるが、それでも従来法と同等以上の性能を示している。

本方式の大きな特長は(1)推論ルールの自動分割(メンバーシップ関数の最適自動決定と(2)自己学習機能、の2点にある。実験結果は(1)に対して、3つの応用事例に関して言えば、従来法でも最適に近い性能を示していると考えられ本法式の劇的な効果は見られない。もっと従来法では困難な複雑なタスクに適用して初めて大きな性能差がでるものと考えられる。

(2)に関してはまだ評価をしていない。本方式の第2ステップ開発として、ニューラルネットの学習機能を用いた動的なファジィ推論規則の実現を今後行う予定である。

## 6. 今後の研究の展開

ニューラルネットの研究は筆者の現在の業務である音声認識の立場からスタートしている。「1. はじめに」で述べたように、この立場でのニューラルネットに対しては、聴覚モデルの構築を通じた聴覚機構の工学的実現・解明と、パターン認識のツールという2面の取組がある。

第1の聴覚モデルに対しては、今後とも息長く地道に研究を続けて音声認識の前処理部として完成させたいと思っている。現在は末梢系のモデルを対象としているが、できれば中枢系にまで拡大し中枢系の処理メカニズムに新しい音声認識のヒントが得られればとも考えている。この場合、聴覚系よりも生理学的研究が進んでいる視覚系の処理メカニズムが大いに参考になろう。

第2のパターン認識のツールとしてニューラルネットを捕える立場からは、性能と限界を明らかにしたいと考えている。このためには色々なタスクに適用しその過程で色々なノウハウを蓄積していく必要があるだろう。ニューラルネットの適用がうまくいくかどうかはそのモデリングにかかっている。このモデルというのはfeed-forward型であるとか、相関型モデルであるとか、ボルツマン型モデルであるといったニューラルネットそのもののことではない。いかなる入力をいかなる規模のモデルに割り付け、いかなる出力を得るかというモデリングである。ニューラルネットは勿論万能ではない。音声波形を入力し認識結果を出すようなモデリングは当然期待できない。結局、そのタスクの物理的・数学的性質に精通した者が最適なモデリングをすることになるだろう。ある程度ニューラルネットの応用実験の経験を積むと、このモデリングの勘処のようなものが掴めてくる。今後ニューラルネットの応用が益々増えてくると思われるが、誰かがこのようなノウハウ的なものを経験しておけば、その初期オフセットの時間を費やすことなくこの技術を利用できる。本解説で述べた応用が多方面に渡っているのは、直接筆者の担当業務だけを考えるのではなく社内への技術利用の普及を考えてのことである。これが、ニューラルネットをツールとして考えた時の今後の筆者等の研究姿勢である。

第3の課題は高速化についてである。音声認識や画像認識のようなタスクの場合、一般にニューラルネットの規模は大きくなり学習に非常な時間がかかる。



例えば、前述したホルマント抽出では、3つのモデルを同時に投入したこともあるが、EWSで1ヶ月近く学習を行なっている。これでも音声データとしてみればまだまだ不足で、もっと大量のデータを投入する必要があるかと考えている。研究段階ではアイデアの試行錯誤を行なうものであり、結果がでるまでの時間が長いと思考との同期が取れない。このためには、学習の高速化が切実なものになってくる。金を惜しみ無く投入しスーパーコンピュータを導入した時の高速効率よりも知恵を出した方がはるかに高速になる。非公式ではあるが、学習ノウハウの積み重ねで100倍ほど高速になったとのウササを、同じような研究者の間でよく耳にする。単なるノウハウではなく数学的なサポートがあればbackpropagationを超える学習アルゴリズムの発見も可能であろう。今後の展開を考えればこの点についても注力していく必要がある。

最後には実現化の問題である。ニューロンチップで対応するのか、並列処理マシンで対処するのか。並列処理マシンでもコネクションマシンのような粗結合のマシンを考えるのか、密結合のマシンで対処するのか。それぞれ毎に考慮していかなければならない課題は多いが、その中であって筆者は密結合のマシンでのニューロ演算について注意を払っている。研究的にはLSI化や話題の数千数万のCPUを結合させることに興味はあるが、ニューロ技術が機器に組み込まれた場合、コスト的な問題と学習なしのニューラルネット演算速度を考えれば、数個または数十個のCPUが他のプログラムと兼用しながら使われるのが現実的もしくは多いのではないかと思うからである。この時の演算を如何に効率よく行なうかを一度検討しておく必要がある。この過程で、ニューロチップアーキテクチャーへのアイデアがでたり、学習アルゴリズムの高速化のアイデアも生まれよう。

筆者らは密結合型30CPU計算機(Sequent社 Symmetry S81)をベースに、この話題も踏まえ、前述の研究姿勢でのぞむつもりである。

## 7. ま と め

最近の筆者のニューラルネットに関わる研究で、応用に関する既発表の研究を概説した。音声処理・文字認識・ファジィの分野への応用を通じて、いかに

モデリングするか、いかにニューラルネットを適用するかの参考になれば幸いである。最後に、ニューラルネットへの取組みに対する4つの方向を示し、筆者等の研究姿勢を述べた。

#### 参考文献

- 1) J. Hutchinson, C. Kock, J. Luo, and C. Mead, "Computing Motion Using Analog and Binary Resistive Networks", IEEE COMPUTER, Vol. 21, No. 3, pp 52-63, Mar. 1988
- 2) 甘利俊一, "神経回路網の数理", 産業図書, 昭53年
- 3) 高木, 坪香, "ニューラルネットを用いた擬似ホルトマン抽出フィルタリング", 音響学会講演論文集 3-P-11, pp 249-250, 昭63年3月
- 4) D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning Representation by Back Propagation Errors", Nature, Vol. 323, pp 533-536, Oct. 9 1986
- 5) 駒木根, 平原, "蝸牛の周波数分析機能を擬似する フィルターバンクの一構成法", 信学技報 SP 87-45, 昭62年7月
- 6) 高木, 原, "ニューラルネットを用いたプロソディ制御—ピッチの制御について—", 音響学会講演論文集 3-P-12, pp 251-252, 昭63年3月
- 7) T. J. Sejnowski and C. R. Rosenberg, "Parallel Networks that Learn to Pronounce English Text", Complex Systems, Vol. 1, pp 145-168, 1987
- 8) 藤崎, 須藤, "日本語単語アクセントの基本周波数パターンとその生成機構のモデル", 音響学会誌, Vol. 27, No.9, pp 445-453, 昭46年
- 9) 高木, 坪香, "ニューラルネットを用いた音韻セグメンテーション", 音響学会講演論文集 2-P-7, pp 215-216, 昭和63年10月
- 10) 金寺登, "音韻セグメンテーションにおけるニューラルネットの利用", 音響学会講演論文集 3-P-14, pp 255-256, 昭63年3月
- 11) 戸川隼人, "マトリクスの数値計算", オーム社, 昭46年
- 12) 香田, 高木, 木 "ニューラルネットによる手書き英数字認識—モデル規模および学習データ数に関する評価検討—", 信学技報 PRU 88-57,

昭 63 年 9 月

- 13) D. J. Burr, "A Neural Network Digit Reconizer" IEEE-SMC, pp 1621-1625, 1986
- 14) 森, 横澤, 梅田, "PDP モデルによる手書き漢字認識", 信学技報, MBE 87-156, Vol.87, No.428, pp 407-414, 昭 63 年 3 月
- 15) 二矢田, 藤井, 森井, "不特定話者を対象とした音声認識法のための動的パラメータと距離尺度に関する考察", 信学論A, Vol. J 69-A, No.5, pp 628-638, 1986
- 16) 林, 高木, "神経回路網モデルによるファジィ推論の定式化", 第 4 回ファジィ・システム・シンポジウム (国際ファジィ学会日本支部), pp 55-60, 昭 63 年 5 月
- 17) 林, 高木, "推論規則を自己獲得するニューラルネット駆動型ファジィ推論", 第 14 回システム・シンポジウム (計測自動制御学会), pp 59-64, 昭 63 年 8 月
- 18) H. Takagi and I. Hayashi, "ARTIFICIAL - NEURAL\_NETWORK-DRIVEN FUZZY REASONING", International Workshop on Fuzzy System Application (IFSA), pp 217-218, Aug. 1988
- 19) 高木, 林, "ニューラルネット駆動型ファジィ推論", 電子情報通信学会秋季全国大会 D-1-169, 昭 63 年 9 月
- 20) 林, 野村, 高木, 長坂, "ニューラルネット駆動型ファジィ推論の提案", 「あいまい知識処理と知的制御システム制御」シンポジウム (計測自動制御学会関西支部), pp 31-36, 昭 63 年 9 月
- 21) 姜, 菅野, "ファジィモデリング", 計測自動制御学会論文集, Vol.23, No.6, pp 650-652, 昭 62 年
- 22) 菅野, "システムのファジィモデリング", システムと制御チュートリアル講座 '87, 「制御工学へのガイドライン」, pp 79-98, 昭 62 年
- 23) 近藤, "モデルの次数を推定する改良型GMDH", 計測自動制御学会論文集, Vol.22, No.9, pp 8-14, 昭 61 年
- 24) S. Fujita and H. koi, "Application of GMDH to Environmental System

Modelling and Management”, SELF-ORGANIZATION METHODS IN MODELING — GMDH Type Algorithms — edited by S. J. Farlow, MARCEL DEKKER, INC., 1984

- 25) W. S. McCulloch and W. Pitts, “A Logical Calculus of the Ideas Imminent in Nervous Activity”, Bulletin of Mathematical Biophysics, Vol. 5, pp 115–133, 1943
- 26) M. Minsky and S. Papert, “Perceptrons”, MIT Press, 1969 (斎藤正男訳“パーセプトロン”, 東京大学出版会, 1971)
- 27) J. J. Hopfield and D. W. Tank, “Neural computation of decisions in optimization problems”, Biol. Cybernetics, Vol. 52, pp 141–152, 1985