論文要旨

網膜から外側膝状体，第 1 視覚野，側頭葉に至る人間の視覚系処理過程をモデル化したニューラルネットワークに TAM(Topographic Attentive Mapping) ネットワークがある．TAM ネットワークは，入力層，基盤層，カテゴリー層，出力層の 4 層からなり，共振学習とビジランス機能を用いて，パターン認識に有効な手法である．しかし，学習データに対して過学習となる傾向があり，ネットワーク構造の表現問題や入力属性の選択問題などが存在する．本論文では，TAM ネットワークのデータ処理過程がファジィ推論過程と類似していることから，情報量を用いた 3 種類のプルーニングルールを定義し，入力層の属性の重要度と 4 層間のリンクの結合の強さを評価して，リンクとノードを削除する新たなプルーニング手法を提案する．また，その有効性を検討するため，数種類の数値例に適用して TAM ネットワークの学習性と頑健性を検討する．

表題：TAM ネットワークのプルーニング手法の提案
著者：林　勲・James R. Williamson

論　文

TAM ネットワークのプルーニング手法の提案*

林　勲† ・ James R. Williamson‡

A Proposal of Pruning Method for TAM Network*

Isao Hayashi† and James R. Williamson‡

The TAM (Topographic Attentive Mapping) Network based on a biologically-motivated neural network model is an especially effective model. When the network makes an incorrect output prediction, the attentional feedback circuit modulates the learning rates and adds a node to the category layer in order to improve the network's prediction accuracy. In this paper, a pruning algorithm for reducing links and nodes at the layers is proposed. The usefulness of the algorithm is also illustrated.

1. はじめに

人間の視覚系では，網膜へ入力された光情報は桿体細胞や錐体細胞の光受容細胞から外側膝状体を介して第 1 視覚野に入力され，コラム構造で方位や方向，色などが識別処理された後，側頭葉で形状が認識される [1]．この視覚処理をモデル化したニューラルネットワーク [2] に，受容野モデル [3]，V1-V2 モデル [4]，ネオコグニトロン [5]，ARTMAP[6,7]，TAM[8,9] などがある．特に，TAM(Topographic Attentive Mapping) は共振学習とビジランス機能を持ち，パターン認識において有効な手法である．TAM ネットワークは，入力層，基盤層，カテゴリー層，出力層の 4 層構造からなる．入力層はガングリオン細胞に対応し，受容野の ON センター OFF サラウンド (OFF センター ON サラウンド) 機能 [10,11] を模擬するため，属性データを分布データとして取り扱う．基盤層は外側膝状体に対応し，視覚野への中継機能と興奮性・抑制性学習を構造化している．出力層は第 1 視覚野あるいは側頭葉を構造化している．与えられた教師値と出力値に差がある場合，基盤層での興奮性学習，出力層から基盤層へのフィードバック信号とビジランスパラメータによる抑制性学習，およびカテゴリー層ノードの増設によるパターンモデルの生成など，パターン問題に対する精度の高い学習が可能である．しかし，カテゴリー層ノードの増加は正のみに限定されているので，学習データに対して過学習となる傾向があり，必ずしも評価データに対して有効な結果が導出できない場合がある．

本論文では，入力層から出力層までの 4 層の各ノードにおいてリンクとノードを削除する新たなプルーニング手法を提案する [12,13]．ニューラルネットワークのプルーニング手法は 3 種類に分類できる [14]．誤差関数に情報量などの評価関数を追加して学習パラメータの更新関数にパラメータ値の衰退関数を用いる方法，各ノードでのリンク方向性ベクトルを計算して類似ベクトルを融合する方法，学習後にノード数やシナプス荷重を増減させその応答値を評価する方法である．本手法では，これらの三つの手法とは異なった方法を提案する．まず，TAM ネットワークのデータ処理過程がファジィ推論過程 [15,16] と類似していることから，ファジィ ID3 での情報量 [17–19] を用いて入力層の属性の重要度を計算し，必要な属性を逐次的に抽出する手法を提案する．カテゴリー層ノードがファジィルールを表現していることから，ファジィルールの重要度も同時に得ることができる．次に，4 層間のリンクの結合の強さを評価し，リンクとノードを削除するための手法を提案する．このリンクの結合の強さを評価するため，占有度を用いた 3 種類のプルーニングルールを定式化した．さらに，本手法の有効性を検討するため，データベースサイトから公開データを入手し，このデータを用いて TAM ネットワークの学

* 原稿受付　2003 年 5 月 20 日
† 阪南大学　大学院　企業情報研究科　Graduate School of Corporate Information, Hannan University; 5-4-33, Amami-higashi, Matsubara, Osaka 580-8502, JAPAN
‡ Lockheed Martin Corp.　Theater Missile Defense, Group 32, MIT, Lexington, MA, U.S.A.
Key Words: neural networks, pruning methods, fuzzy rules, knowledge acqusition.

習性とプルーニング手法の頑健性を検討する．本手法はファジィID3の情報量を用いて学習後にノード数の占有度を計算してネットワークの出力値を評価していることから，プルーニング手法の分類の第 1 手法と第 3 手法との融合型ともいえる．ここでは，まず TAM ネットワークについて概説し，次に，TAM ネットワークのプルーニング手法の定式化を行ない，最後に，数値例に適用してプルーニング機能の有効性について議論する．

2. TAM ネットワーク

TAM ネットワークは Williamson[8] によって提案された視覚系ニューラルネットワークである．構造を Fig. 1 に示す．下位層から上位層に向かって，入力層，基盤層，カテゴリー層，出力層の 4 層からなる．

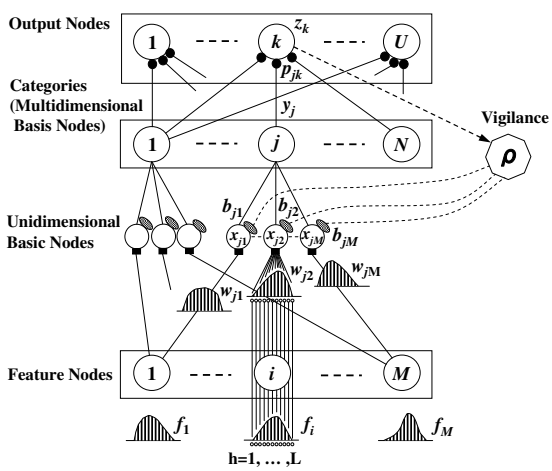


Fig. 1 TAM network

入力層は網膜細胞の受容野を想定しており，入力信号を分布データとして受信する．いま，与えられたデータ集合 D が M 個の入力属性と 1 個の出力属性からなる R 個のデータとする．第 i 番目の入力属性における第 s 番目のデータを $f_{si}, s=1,2,\cdots,R$ で表し，その出力値を k とする．まず，f_{si} をその大きさにより順位づけし，$1,2,\cdots,s,\cdots,R$ とする．次に，ランク法を用いて各属性ごとに正規化する．

$$I_{si}=\frac{s-0.5}{R}, \qquad i=1,2,\cdots,M \tag{1}$$

入力層では，この I_{si} を用いて受容野での信号入力を分布データ f_{sih} として受信する．

$$f_{sih}=\frac{exp[-0.5(LI_{si}-h+0.5)^2]}{\sum_{h'=1}^{L} exp[-0.5(LI_{si}-h'+0.5)^2]} \tag{2}$$

ただし，h は分布の離散量 $h=1,2,\cdots,L$ を表す．なお，入力層への入力信号はデータ集合 D から 1 個ずつ逐次的に入力されるので，以後，簡素化のため，入力 f_{sih} を f_{ih} として表記する．

基盤層とカテゴリー層は神経節細胞もしくは外側膝状体に対応し，視覚野への中継機能と興奮性，抑制性学習を構造化している．基盤層では，シナプス荷重 w_{jih} により j 番目のノードでの活性値 x_{ji} を計算し，第 1 視覚野の出力層から選択的注意のフィードバック信号を受ける．また，カテゴリー層では，基盤層の信号を統合し，出力層への出力 y_j を計算する．

$$x_{ji}=\frac{\sum_{h=1}^{L} f_{ih}w_{jih}}{1+\rho^2 b_{ji}} \tag{3}$$

$$y_j=\prod_{i=1}^{M} x_{ji} \tag{4}$$

ただし，b_{ji} は抑制のためのシナプス荷重であり，ρ はフィードバック信号を表すビジランスパラメータである．後述の学習モードにおいて，w_{jih} は入力 f_{ih} に近似するように学習されるので，カテゴリー層の各ノードは入力層の各入力パターンを記憶した構造となる．

出力層は第 1 視覚野を構造化している．k 番目の出力ノードでは，j 番目のカテゴリーノードとのシナプス荷重 p_{jk} を介して出力値 z_k を計算する．最大値を有するノード番号 K が TAM ネットワークの出力値として出力される．

$$z_k=\sum_{j=1}^{N} y_j p_{jk} \tag{5}$$

$$K=\{k|\max_k z_k\} \tag{6}$$

いま，K^* を正しい教師値としよう．TAM ネットワークの出力値 K が K^* と一致しない場合には，ビジランスパラメータ ρ を初期値 $\rho=0$ から $\rho^{(step)}$ 分だけ上昇させ，b_{ji} とともにカテゴリー層のノードの活性値 x_{ji} を抑制し，相対的に他のノードの活性値を上昇させる．これは網膜に投影された対象への注意をより強調させる第 1 視覚野から外側膝状体への “選択的注意” の発動機能を想定している．

ρ は $z_{K^*}/z_K \geq OC$ の条件が満足されるか最大値 $\rho^{(max)}$ になるまで上昇を続ける．ただし，OC は評価のしきい値であり，$0 \leq OC \leq 1$ である．ビジランスパラメータ ρ が最大値 $\rho^{(max)}$ を越えた場合には，現在のカテゴリー層のノード構成では K^* を十分に表現できないと考え，カテゴリー層のノードを新たに 1 個分増加させる．

$$\begin{aligned}&\text{If } z_{K^*}/z_K < OC \text{ then repeat}\\ &\quad (a)\ \rho=\rho+\rho^{(step)}\\ &\quad (b)\ \text{equation } (3)-(6)\\ &\text{until either } z_{K^*}/z_K \geq OC \text{ or } \rho \geq \rho^{(max)}\end{aligned} \tag{7}$$

一方，$z_{K^*}/z_K \geq OC$ を満足する場合には，現在のネットワーク構造で K^* を正しく表現していると考え，出力

値 K をより K^* に近似させるため，学習モードに入る．

学習モードでは，まず出力層からカテゴリー層へのフィードバック信号 y_j^* を計算する．

$$z_k^* = \begin{cases} 1\ ;\ \text{if } k=K^* \\ 0\ ;\ \text{otherwise} \end{cases} \tag{8}$$

$$y_j^* = \frac{\prod_{i=1}^{M} x_{ji} \times \sum_{k=1}^{U} z_k^* p_{jk}}{\sum_{j'=1}^{N} \prod_{i=1}^{M} x_{j'i} \times \sum_{k=1}^{U} z_k^* p_{j'k}} \tag{9}$$

シナプス荷重 b_{ji}, p_{jk}, w_{jih} は次式により更新される．

$$\triangle b_{ji} = b_j^{(rate)} y_j^* (x_{ji} - b_{ji}) \tag{10}$$

$$\begin{aligned} \triangle p_{jk} &= p_j^{(rate)} y_j^* (z_k^* - p_{jk}) \\ &= \frac{\alpha y_j^* (z_k^* - p_{jk})}{\alpha + n_j} \end{aligned} \tag{11}$$

$$\begin{aligned} \triangle w_{jih} &= w_j^{(rate)} y_j^* (f_{ih} - w_{jih}) \\ &= \frac{\alpha y_j^* (f_{ih} - w_{jih})}{\alpha\beta(M) + n_j} \end{aligned} \tag{12}$$

ただし，

$$\beta(M) = \frac{\lambda^{1/M}}{1-\lambda^{1/M}}, \qquad \lambda \in (0,1) \tag{13}$$

$$\triangle n_j = \alpha y_j^* (1 - n_j) \tag{14}$$

であり，α と λ は係数である．なお，$b_j^{(rate)}, p_j^{(rate)}, w_j^{(rate)}$ は学習係数であり，$b_j^{(rate)}$ は定数，$p_j^{(rate)}$ はシミュレーティッドアニーリング法の補正値，$w_j^{(rate)}$ はさらに入力次元 M による補正項 $\beta(M)$ を加えた係数である．

学習モードでは，フィードバック信号 y_j^* を介して，w_{jih} は f_{ih} に近似するように学習される．また，p_{jk} は z_k^* に近似するように学習され，b_{ji} は x_{ji} に近似するように学習される．これらの学習は Grossberg の winner 学習 [1] を用いており，共振学習 [6] という．

なお，学習は TAM ネットワークに入力データ f_{ih} を逐次的に投入するごとに行い，全学習データはデータ集合 D を 1 エポックとして数エポック回数分とする．

3. プルーニング手法の提案

TAM ネットワークは学習性に優れているが，カテゴリー層ノードは単調増加的に付加されるので，過学習となる傾向がある．また，ネットワークの構造を明示的に表現できず，入力属性の削減問題も存在する．これらの問題を解決する一手法として，ここではプルーニング手法を提案する．

いま，TAM ネットワークの構造を議論するため，4 層を入力層・基盤層・カテゴリー層の下位層とカテゴリー層・出力層の上位層との二つに分割しよう．下位層では，基盤層ノードのシナプス荷重 w_{ji} は入力層の入力 f_i に近似するように学習されるので，カテゴリー層の各ノードは，全学習データの入力パターンをその種別ごとに記憶している．一方，上位層では，出力層ノードとカテゴリー層ノードとがシナプス荷重 p_{jk} を介して結合している．下位層の構造から，この p_{jk} はカテゴリー層ノードでの出力パターンに対する度合い，すなわち，全学習データの入力パターンに対する出力パターンの度合いを表現していると考えられる．この関係を次のルール表現を用いて記述しよう．

$$\left.\begin{array}{ll} r_1: & \text{if } f_1 \text{ is } w_{11} \text{ and } \cdots \text{ and } f_M \text{ is } w_{1M} \\ & \text{then } C_1 = p_{11}, \cdots, C_U = p_{1U} \\ \vdots & \vdots \qquad\qquad\qquad \vdots \\ r_j: & \text{if } f_1 \text{ is } w_{j1} \text{ and } \cdots \text{ and } f_M \text{ is } w_{jM} \\ & \text{then } C_1 = p_{j1}, \cdots, C_U = p_{jU} \\ \vdots & \vdots \qquad\qquad\qquad \vdots \\ r_N: & \text{if } f_1 \text{ is } w_{N1} \text{ and } \cdots \text{ and } f_M \text{ is } w_{NM} \\ & \text{then } C_1 = p_{N1}, \cdots, C_U = p_{NU} \end{array}\right\} \tag{15}$$

ただし，$C_k, k=1,2,\cdots,U$ は出力のクラスであり，$r_j, j=1,2,\cdots,N$ はルール番号を表す．

(15) 式のルールの表現形式から，いま，w_{ji} をファジィ集合のメンバシップ関数，p_{jk} を後件部の実数値と解釈すれば，(15) 式はファジィルールと解釈することができる．このことは，本論文で提案するプルーニング機能により，TAM ネットワークから不必要な w_{ji} と p_{jk}，およびカテゴリー層ノード r_j が削除されるとすると，全ファジィルールから不必要なファジィルールがとり除かれ，TAM ネットワークの構造がファジィルールとして抽出されることを示す．したがって，本手法はニューラルネットワークの構造を明示的に表現するファジィニューラルネットワーク [16] の一手法であるといえる．

与えられたデータ集合をファジィルールとして獲得する手法にファジィ ID3 がある．ファジィ ID3 では，属性ごとにデータ情報量を求め，情報量が最大となる属性を逐次抽出して，全データ集合を表現するファジィルールを得る．この情報量基準を TAM ネットワークのプルーニング手法に応用する．具体的には，情報量が最大となる属性を逐次求め，全データを表現するリンクとカテゴリー層ノードを得て，それ以外のリンクとノードを削除する．

いま，情報量最大化により得られた属性の集合を I^* とする．属性集合 I^* に対する (4) 式の活性値の統合量は次のように計算できる．

$$\delta_{js} = \prod_{i' \in I^*} x_{ji's} \tag{16}$$

ただし，(15) 式より，この δ_{js} はファジィルールの前件部の統合メンバシップ値を表していることがわかる．

ファジィ ID3 では，条件付き情報量を計算するために，属性ごとのクラス k の占有率を計算する．i 番目の属性を I^* に加えることによるデータ集合 D におけるクラス k の占有率は次のように計算できる．

$$G_{jk}=\frac{\sum_{s'\in\psi_k}\gamma_{js'}\times p_{jk}}{\sum_{s=1}^{R}\gamma_{js}\times p_{jk}} \tag{17}$$

ただし，

$$\gamma_{js}=x_{jis}\times\delta_{js} \tag{18}$$

であり，ψ_k はデータ集合 D でのクラス k をもつ入力データの集合を表す．

(16) 式 ~ (18) 式により，i 番目の属性を選択することによる条件付き情報量 $H(i)$ を次のように計算する．

$$H(i)=-\sum_{j=1}^{N}g_j\sum_{k=1}^{U}G_{jk}\log_2 G_{jk} \tag{19}$$

ただし，

$$g_j=\frac{\sum_{s=1}^{R}x_{jis}}{\sum_{j=1}^{N}\sum_{s=1}^{R}x_{jis}} \tag{20}$$

であり，g_j はカテゴリー層ノードに対する事前確率である．

選択される i^* 番目の属性は情報量最大値をもつ属性である．

$$i^*=\{i|\max_i H(i)\} \tag{21}$$

データ表現のために必要な属性は (21) 式によって逐次的に選択されていくので，その選択順位は属性の重要度を表していると考えられる．この属性の選択過程において，次の三つの観点から各リンクとノードとを評価し，不必要なリンクとノードを削除する．

- 第 j 番目のカテゴリー層ノードにおいて，クラス層ノードとの結合の強さを評価し，不要な結合を削除する．
- 第 j 番目のカテゴリー層ノードにおいて，入力層の各属性との結合の強さを評価し，不要な結合を削除する．
- 第 k 番目のクラス層ノードにおいて，カテゴリー層ノードとの結合の強さを評価し，不要な結合を削除する．

これらを次の三つのプルーニングルールとして定式化する．

【プルーニングルール 1】 第 j 番目のカテゴリー層ノードにおいて，第 k 番目のクラス層ノードとの間で次の条件を満足した場合，第 j 番目のカテゴリー層ノードと第 $k',k'=1,2,\cdots,U,k'\neq k$ 番目のクラス層ノードとの結合を削除する．同時に，第 j 番目のカテゴリー層ノードと第 $i'\notin I^*$ 番目の属性との結合を削除する．

$$G_{jk}\geq\eta \tag{22}$$

ただし，η はしきい値である．

【性質 1】 続くプルーニング過程で，$\sum_{s\in\psi_k}x_{ji''s}\geq\sum_{s\notin\psi_k}x_{ji''s}$ を満足する第 i'' 番目の属性が選択された場合には，(22) 式は満足される．

【プルーニングルール 2】 第 j 番目のカテゴリー層ノードにおいて，次の条件を満足した場合，第 j 番目のカテゴリー層ノードと第 i 番目および第 $i'\notin I^*$ 番目の属性との結合を削除する．

$$\frac{1}{R}\sum_{s=1}^{R}\gamma_{js}<\theta \tag{23}$$

ただし，θ はしきい値である．

【性質 2】 (23) 式が一度満足されると，続くプルーニング過程 I'^* でも，$x_{jis}\leq 1$ から $\prod_{i\in I^*}x_{jis}\geq\prod_{i\in I'^*}x_{jis}$ となり，(23) 式の条件は満足される．

【プルーニングルール 3】 クラス K において，次の条件を満足した場合，第 K 番目のクラスと第 $j'\neq j$ 番目のカテゴリーとの結合を削除する．

$$\varphi_{jK}=\frac{\sum_{s\in\Gamma_K}\gamma_{js}\times p_{jK}}{\sum_{j=1}^{N}\sum_{s\in\Gamma_K}\gamma_{js}\times p_{jK}}\geq\xi \tag{24}$$

ただし，$\Gamma_K=\{s|\,K=K^*,K=\max_k\sum_{j=1}^{N}y_j p_{jk}\}$ であり，ξ はしきい値である．

【性質 3】 ξ が次式を満足する場合には，出力値と教師値との一致度は維持される．

$$\xi\geq\max_{k\neq K}\frac{\sum_{s\in\Gamma_K}\sum_{j=1}^{N}p_{jk}}{\sum_{s\in\Gamma_K}\sum_{j=1}^{N}y_{js}p_{jk}} \tag{25}$$

性質 3 について説明しよう．いま，$\phi_k=\{j|\varphi_{jk}\geq\xi\}$ として，クラス k と $j\notin\phi_k$ を満足する j との結合を削除することを考える．他のクラス $k'\neq K$ との関係から，$s\in\Gamma_k$ なる任意の s に対して，次式が満足されなければならない．

$$\sum_{j=1}^{N}\gamma_{js}p_{jk}-\sum_{j\in\phi_k}\gamma_{js}p_{jk}-\max_{k'}\sum_{j=1}^{N}\gamma_{js}p_{jk'}\geq 0 \tag{26}$$

しかし，上式は求められないので，ここでは，次の条件を考える．

$$\sum_{s\in\Gamma_k}\sum_{j\in\phi_k}\gamma_{js}p_{jk}-\max_{k'}\sum_{s\in\Gamma_k}\sum_{j=1}^{N}\gamma_{js}p_{jk'}\geq 0 \tag{27}$$

この条件を満足させるには，

$$\frac{\sum_{s\in\Gamma_k}\sum_{j\in\phi_k}\gamma_{js}p_{jk}}{\sum_{s\in\Gamma_k}\sum_{j=1}^{N}\gamma_{js}p_{jk}}\geq\xi\geq\max_{k'}\frac{\sum_{s\in\Gamma_k}\sum_{j=1}^{N}\gamma_{js}p_{jk'}}{\sum_{s\in\Gamma_k}\sum_{j=1}^{N}\gamma_{js}p_{jk}} \tag{28}$$

となる ξ を決定すれば良い．しかし，ϕ_k はプルーニング過程で決定されるので，γ_{js} は得られない．そのため，$x_{jis} \leq 1$ より，

$$\xi \geq \max_{k'} \frac{\sum_{s\in\Gamma_k}\sum_{j=1}^{N} p_{jk'}}{\sum_{s\in\Gamma_k}\sum_{j=1}^{N}\gamma_{js}p_{jk}} \geq \max_{k'}\frac{\sum_{s\in\Gamma_k}\sum_{j=1}^{N}\gamma_{js}p_{jk'}}{\sum_{s\in\Gamma_k}\sum_{j=1}^{N}\gamma_{js}p_{jk}} \quad (29)$$

とする．以上より，(25) 式が得られた．

プルーニング機能を組みこんだ TAM ネットワークの学習アルゴリズムを以下に示す．

Step 1 データ集合 D の R 個のデータを学習用に用いるための学習データ (TRD: R_T 個) と評価用に用いるための評価データ (CHD: R_C 個) に分割する．ここで，$R_T + R_C = R$ である．R_T 個の学習データを 1 エポックとし，学習のための全データを数エポック回数分とする．

Step 2 学習データを 1 個ずつ逐次的に TAM ネットワークに投入し，(1) 式 ~ (6) 式より TAM ネットワークの出力値 K を得る．

Step 3 $z_{K^*}/z_K < OC$ の場合，$\rho = \rho + \rho^{(step)}$ とする．ρ が最大値になった場合，カテゴリー層のノードを 1 個分増加させる．

Step 4 $z_{K^*}/z_K \geq OC$ を満足する場合，Step6 の学習モードに入る．

Step 5 すべての学習データが終了するまで，Step2 から Step4 までを繰り返す．

Step 6 学習では，(8) 式 ~ (14) 式を用いて，シナプス荷重 b_{ji}, p_{jk}, w_{jih} を変更する．

Step 7 学習が終了した後，プルーニングモードに入る．学習データに対して，(19) 式を用いて第 i 番目の属性の情報量 $H(i)$ を計算する．

Step 8 (21) 式の i^* を選択し，集合 $I^* = \{i^*\}$ を構成する．

Step 9 評価データに対して，プルーニングルール 1 を適用する．

Step 10 評価データに対して，プルーニングルール 2 を適用する．

Step 11 評価データに対して，プルーニングルール 3 を適用する．

Step 12 全結合が削除されたノードを削除する．

Step 13 Step8 において，すべての属性が選択されるまで，Step7 から Step12 までを繰り返す．

アルゴリズム終了後，I^* に登録された属性の順序によって属性の重要度が得られる．また，(24) 式の φ_{jk} は第 k 番目の出力層ノードにおける第 j 番目のカテゴリー層ノードの占有率を計算しているので，出力層のそれぞれのノードにおいてカテゴリー層ノードの重要度が得られる．プルーニング過程終了後，不必要なリンクやノードが削除されたネットワークの構造が得られ，ルール数が調整されたファジィルールが獲得できる．

4. 数値例

ここでは，DELVE(Data for Evaluating Learning in Valid Experiments)[20] のデータを用いて，TAM ネットワークの有効性を検討する．まず，学習データに対する精度を議論するため，画像データの判別問題に TAM ネットワークを適用した．データ集合は 3 × 3 ピクセルからなる 2310 個のデータである．学習データの精度を確認するので，学習データと評価データを同一にして無作為に 70 個を抽出した．入力属性は RGB の色度や輝度などの 16 属性であり，画像の種類 (煉瓦，空，葉，セメント，窓，道，芝) が出力クラスである．学習のためのパラメータ値を次に示す．

エポック数	$= 1$	L	$= 10$
OC	$= 0.8$	$\rho\ step$	$= 0.1$
$\rho\ max$	$= 10.0$	$b_j^{(rate)}$	$= 0.01$
α	$= 0.01$	λ	$= 0.33$
$(Case\,1)$			
η	$= 0.8$	θ	$= 0.01$
ξ	$= 0.8$	条件	$= AND$
$(Case\,2)$			
η	$= 0.3$	θ	$= 0.1E-15$
ξ	$= 0.7$	条件	$= OR$

実験では，条件項目を変えた $Case\,1$ と $Case\,2$ の 2 種類を用意した．ここで，条件 $= AND$ または 条件 $= OR$ とは，プルーニングルール 1 ~ 3 の接合条件を表す．条件 $= AND$ の場合には，カテゴリー層ノードと属性との結合を削除する条件はプルーニングルール 1 と 2 の両条件を満足する場合のみとし，カテゴリー層ノードとクラス層ノードとの結合の削除はプルーニングルール 1 と 3 の両条件を満足する場合のみとする．一方，条件 $= OR$ の場合には，結合の削除はプルーニングルール 1 ~ 3 のいずれかが満足された場合とする．結果の正解率を Table 1 に示す．TAM ネットワークの結果は 15 回の正解率の平均値であり，最も良い結果も同時に示す．**Cart**[21] は決定木の構成手法であり，**1NN** と **KNN-Class**[22] はユークリッド距離に基づいたクラスタリング手法，**ME** と **HME**[23] は混合エキスパートアルゴリズムである．TAM ネットワークの正解率は他手法と比較して同等かもしくは高く，また，プルーニング機能により 16 属性が 8 属性に削除されている．

次に，評価データの頑健性を評価するため，タイタニック号の水難事故データ [20] を用いた判別問題に適用した．タイタニック号の救助された人と救助されなかった人とが判別されている全乗客 2201 名のデータ集合が

Table 1 Result of image data

Methods	Correct Rate [%]	Categories	Features
TAM(Best)	74.3	30	16
TAM+Pruning(Best)	82.9	35	8
TAM	68.1	30	16
TAM+Pruning(Case1)	71.0	33.9	16
TAM+Pruning(Case2)	67.4	33.9	8
Cart	63.9	–	–
1NN	66.2	–	–
KNN-Class	64.6	–	–
ME-EL	69.1	–	–
ME-ESE	68.0	–	–
HME-EL	68.9	–	–
HME-ESE	68.1	–	–
HME-GROW	68.3	–	–

ある．入出力値は，船室クラス (1 等船室，2 等船室，3 等船室，乗務員)，年齢区別 (大人，子供)，性別 (男性，女性) であり，出力値は救助の有無である．全データに対して，乱数で 20 個，40 個，80 個，160 個を抽出して，学習データと評価データを構成した．学習パラメータの値を次に示す．

エポック数	$= 1$	L	$= 5$
OC	$= 0.8$	$\rho\ step$	$= 0.1$
$\rho\ max$	$= 100.0$	$b_j^{(rate)}$	$= 0.01$
α	$= 0.00001$	λ	$= 0.33$
η	$= 0.8$	θ	$= 0.01$
ξ	$= 0.8$	条件	$= OR$

学習データと評価データの正解率を Table 2 に示す．TAM ネットワークの結果は 15 回の正解率の平均値である．TAM ネットワークは学習性に優れ，評価データに対する正解率も他手法よりも良い．プルーニング機能を用いることによりさらに正解率は向上した．プルーニング機能により過学習が押さえられたと解釈できる．

最後に，図形データを用いて変数選択性とファジィルール獲得の有用性を検討した．いま，Fig. 2 と Fig. 3 のような白黒点を判別する図形データが与えられている．この図形データを学習データとして白黒点の判別曲線を表すファジィルールを獲得した．学習パラメータの値を次に示す．

エポック数	$= 10$	L	$= 10$
OC	$= 0.8$	$\rho\ step$	$= 0.1$
$\rho\ max$	$= 100.0$	$b_j^{(rate)}$	$= 0.01$
α	$= 0.0000001$	λ	$= 0.33$

Table 2 Results of Titanic data

	Methods	20	40	80	160
TRD	TAM	83.3	77.5	75.4	74.4
	TAM+Pruning	80.0	87.5	80.0	76.9
CHD	TAM	72.9	75.6	77.5	78.2
	TAM+Pruning	80.0	77.5	78.8	78.8
	Cart	68.4	73.8	76.6	75.3
	1NN	68.7	67.8	68.8	69.5
	KNN-Class	71.6	73.1	74.2	75.7

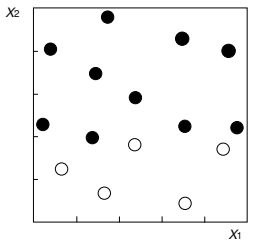


Fig. 2 Training data 1

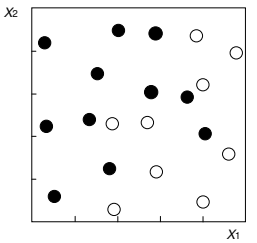


Fig. 3 Training data 2

まず，Fig. 2 を学習データとして TAM ネットワークの学習を行った．学習データに対する正解率は 97.8% となった．入力層，カテゴリー層，出力層の各ノード数は，それぞれ，2，4，2 であった．Fig. 4 の評価データに対して，TAM ネットワークの白黒点を判別する境界線を描いた (Fig. 4 の網掛け部分)．$f_1 = 1$ と $f_1 = 9$ 付近に判別の欠損が認められる．評価データに対する正解率は 82.2% となった．次に，$\eta = 0.005,\ 0.2,\ 0.5$，$\theta = 0.0005,\ 0.005,\ 0.05$，$\xi = 0.1,\ 0.4,\ 0.8$，条件 $= OR$ として，プルーニングアルゴリズムを適用し，評価データに対する正解率を求めた．結果を Table 3 に示す．正解率は 3 回の試行の平均値であり，各ノード数 (入力層− カテゴリー層− 出力層) も同時に表す．パラメータの値によっては正解率が改善していることがわかる．$\eta = 0.005$, $\theta = 0.005$, $\xi = 0.4$ の場合の出力の境界線を Fig. 5 に示す．また，学習回数に対する学習データと評価データの正解率の変化を Fig. 6 に示し，ネットワーク構造を Fig. 7 に示す．属性 f_1 と第 1 番目と第 2 番目のカテゴリー層ノード，およびそれらの結合が削除され，カテゴリー層のノード数から 3 個のファジィルールが抽出されることがわかる．

同様に，Fig. 3 を学習データとして TAM ネットワークの学習を行った．学習データに対する正解率は 81.7% であり，評価データに対する正解率は 71.7% であった．入力層，カテゴリー層，出力層の各ノード数はそれぞれ，2，10，2 となった．次に，プルーニングアルゴリズムを適用した．評価データに対する正解率を Table 4 に示す．パラメータの値を変更しても正解率は改善しなかっ

た．また，ノード数も変化しなかった．これは Fig. 3 の判別線が複雑な非線形曲線であることから，削除すべき TAM ネットワークのノードやリンクが存在しないと考えられる．また，カテゴリー層のノード数はファジィルールの個数を表しているので，非線形な判別線を表現するのに多くのルールが必要であると考えられる．

5. おわりに

TAM ネットワークのプルーニング手法を定式化し，その有効性を議論した．今後，プルーニング後の精度をより向上させるため，η, θ, ξ のパラメータの最適値を得る手法を検討する必要がある．本研究の一部は，文部科学省科学研究補助金（基盤研究 C）課題番号 14580433 の一環として行われた．

Table 3 Results of checking data 1

η	θ \ ξ	0.1		0.4		0.8	
	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
0.005	0.005	73.3	1-2-2	97.8	1-2-2	75.5	1-2-2
	0.0005	91.1	1-3.7-2	84.4	1-3.3-2	86.7	1-4-2
	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
0.2	0.005	73.3	1-2-2	95.5	1-2-2	75.5	1-2-2
	0.0005	93.3	1-3.7-2	88.9	1-3.3-2	88.9	1-4-2
	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
0.5	0.005	73.3	1-2-2	97.8	1-2-2	75.5	1-2-2
	0.0005	91.1	1-3.7-2	84.4	1-3.3-2	86.7	1-4-2

Table 4 Results of checking data 2

η	θ \ ξ	0.1		0.4		0.8	
	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
0.005	0.005	63.3	1-9.3-2	58.3	1-9.7-2	58.3	1-9.7-2
	0.0005	71.7	2-10-2	71.7	2-10-2	71.7	2-10-2
	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
0.2	0.005	63.3	1-9.3-2	56.7	1-9.7-2	56.7	1-9.7-2
	0.0005	71.7	2-10-2	71.7	2-10-2	71.7	2-10-2
	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
0.5	0.005	63.3	1-9.3-2	58.3	1-9.7-2	58.3	1-9.7-2
	0.0005	71.7	2-10-2	71.7	2-10-2	71.7	2-10-2

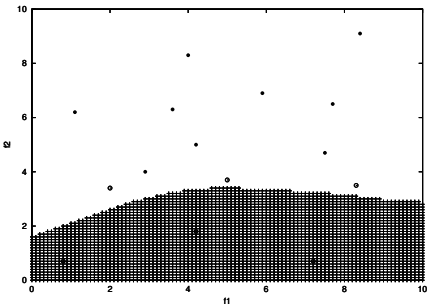


Fig. 4 TAM network

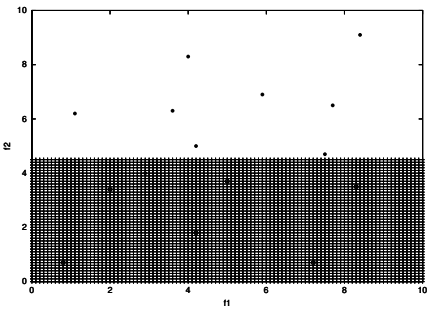


Fig. 5 TAM network with pruning

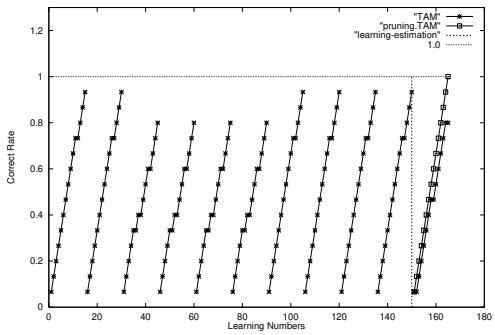


Fig. 6 Correct rate

参 考 文 献

[1] 松村: ニューロサイエンス入門, サイエンス社 (1995)
[2] S.Grossberg: How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex; *Spatial Vision*, Vol. 12, No. 2, pp. 163–185 (1999)
[3] 岡島，今岡: 情報量最大化と生体視覚細胞の受容野; 電子情報通信学会論文誌, Vol. J83-A, No. 6, pp. 620–628 (2000)
[4] H.Neumann and W.Sepp: Recurrent V1-V2 interaction in early visual boundary processing; *Biological Cybernetics*, Vol. 81, pp. 425–444 (1999)
[5] K.Fukushima: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position; *Biological Cybernetics*, Vol. 36, pp. 193–202 (1980)
[6] G.A.Carpenter, S.Grossberg and J.Reynolds: ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network; *Neural Networks*, Vol. 4, pp. 565–588 (1991)
[7] G.A.Carpenter, S.Grossberg, N.Markuzon,

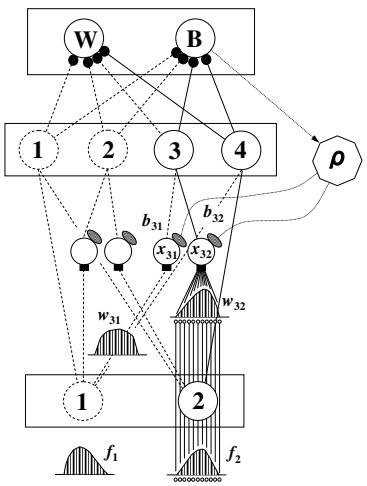


Fig. 7 TAM network structure after pruning

J.Reynolds and D.B.Rosen: Fuzzy ARTMAP: A neural network architecture for incremental learning of analog multidimensional maps; *Neural Networks*, Vol. 3, pp. 698–713 (1992)

[8] J.R.Williamson: Self-organization of topographic mixture networks using attentional feedback; *Technical Report CAS/CNS*, Vol. TR-99-0273 (1999)

[9] J.R.Williamson: Self-organization of topographic mixture networks using attentional feedback; *Neural Computation*, Vol. 13, pp. 563–593 (2001)

[10] 松田: 知覚心理学の基礎, 培風館 (2000)

[11] S.Marcelja: Mathematical description of the responses of simple cortical cells; *J. Opt. Soc. Am.*, Vol. 70, No. 11, pp. 1297–1300 (1980)

[12] 林，前田: TAM Network のブルーニング手法の一提案; 第 *16* 回ファジィシステムシンポジウム講演論文集, pp. 27–130 (2000)

[13] I.Hayashi and J.R.Williamson: Acquisition of fuzzy knowledge from topographic mixture networks with attentional feedback; *the International Joint Conference on Neural Networks (IJCNN'01)*, pp. 1386–1391 (2001)

[14] R.Reed: Pruning algorithms - A survey; *IEEE Transactions on Neural Networks*, Vol. 4, pp. 740–747 (1993)

[15] B.Kosko: *Neural networks and fuzzy systems, A dynamical approach to machine intelligence*, Prentice-Hall (1992)

[16] 林，馬野: ファジィ・ニューラルネットワークの現状と展望; 日本ファジィ学会誌, Vol. 5, No. 2, pp. 178–190 (1993)

[17] M.Umano, H.Okamoto, I.Hatono, H.Tamura, F.Kawachi, S.Umezu and J.Kinoshita: Fuzzy decision trees by fuzzy ID3 algorithm and its applications to diagnosis system; *Proceeding of Third IEEE International Conference on Fuzzy Systems*, Vol. 3, pp. 2113–2118 (1994)

[18] I.Hayashi: Acquisition of fuzzy rules using fuzzy ID3 with ability of learning for AND/OR operators; *Proceeding of 1996 Australian New Zealand Conference on Intelligent Information Systems*, pp. 187–190 (1996)

[19] 林，前田，小澤: AND 演算子を学習するファジィ ID3 の一提案; 日本ファジィ学会誌, Vol. 11, No. 4, pp. 677–682 (1999)

[20] C.E.Rasmussen, R.M.Neal, G.E.Hinton, D var Camp, M.Revow, Z.Chahramani, R.Kustra and R.Tibshirani: The delve manual, version 1.1; http://www.cs.toronto.edu/~delve (1996)

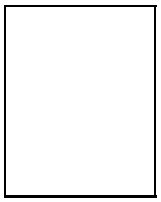
[21] L.Breiman, J.H.Friedman, R.A.Olshen and C.J.Stone: Classification and regression trees; *Wadsworth International Group* (1984)

[22] S.R.Waterhouse, D.J.C.Mackay and A.J.Robinson: Bayesian methods for mixtures of experts; *Advances in Neural Information Processing Systems* (D.S.Touretzky, M.C.Mozer and M.E.Hasselmo Eds.), Morgan Kauffman (1996)

[23] J.A.Robert, M.I.Jordan, S.J.Nowlan and G.E.Hinton: Adaptive mixtures of local experts; *Neural Computation*, Vol. 3, No. 1, pp. 79–87 (1991)

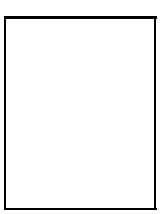
著 者 略 歴

林 勲 (正会員)



1957 年 10 月 11 日生．1981 年 3 月大阪府立大学工学部経営工学科卒業後，シャープ (株) 入社．1985 年 3 月大阪府立大学大学院工学研究科経営工学専攻博士前期課程修了．松下電器産業 (株) 中央研究所を経て，1993 年 4 月阪南大学商学部経営情報学科講師，1997 年 11 月経営情報学部教授，2000 年 4 月大学院企業情報研究科教授となり現在に至る．その間，1999 年 4 月 ~ 2000 年 3 月ボストン大学客員研究員．視覚系ニューラルネットワークとファジィ知識獲得の研究に従事．工学博士．1991 年 4 月電気関係学会関西支部連合大会講演会奨励賞受賞．日本知能情報ファジィ学会，日本神経回路学会，日本基礎心理学会，日本視覚学会，IEEE，INNS などの各会員．

James R. Williamson



1995 年 8 月ボストン大学大学院 Cognitive and Neural Systems 研究科博士課程修了．1997 年 10 月エンキド・リサーチ (株) を経て，1998 年 9 月ボストン大学大学院 Cognitive and Neural Systems 研究科助手，2000 年 9 月イント・ネットワーク (株) 入社，2001 年 10 月ロッキード・マーチン (株) 入社，上級研究員，現在に至る．視覚モデルとニューラルネットワークの研究に従事．工学博士．IEEE, Neural Computation, Cognitive Neuroscience, Information Sciences などの各会員．