Acquisition of Fuzzy Knowledge from Topographic Mixture Networks with Attentional Feedback

Isao Hayashi *, James R. Williamson **
* Hannan University
5-4-33, Amami-higashi, Matsubara, Osaka 580-8502, Japan, email: ihaya@hannan-u.ac.jp
** Into Networks
150 Cambridgepark Drive, Cambridge, MA 02140, email: jamesw@intonet.com

Abstract

The TAM Network (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 method for reducing the number of category and feature nodes is formulated. We discuss here the formulation and show its usefulness through some examples.

1 Introduction

In the human visual system, an image on retina enters to the lateral geniculate nucleus(LGN) through rod cells and pyramidal cells, and is sent to the primary visual cortex, V1, and V2. In the V1, the contour, color and texture of the image are detected, and the contour direction is recognized. The information is divided into two groups whose paths are for the temporal lobe and the vertex lobe[1]. The models based on the visual system are BCS[2], FCS[2], Neocognitron[3], Recurrent V1-V2 model[4], ARTMAP[5], fuzzy ARTMAP[6], Gaussian ARTMAP[7] and TAM[8].

The topographic attentive mapping (TAM) network is especially a useful model, which provides resonance learning and vigilance mechanism inside. The TAM network has three layers constructed by feature, category and class layers. The feature layer is constructed to imitate the retina, and the feature node has a receptive field. The category layer imitates the LGN. The supervised output are given to the class layer imitated the V1. If there is the error between the supervised output and the output of TAM network, the "attention" is invoked and the vigilance parameter is raising. When the error is enough to be small, the learning step starts. The TAM network has a powerful performance. However, the incorrectness by overlearing and/or overfitting might occur for checking data since the nodes at the category layer are only added.

In this paper, a pruning method of the TAM network are proposed, and fuzzy rules are extracted from the network structure. The plenty of pruning methods have been proposed[9], e.g., the method added terms to the objective function and the method which estimates the sensitivity of the error function. In this paper, an information entropy[10][11] for estimating the strength of the link connections is formulated, and three pruning rules are defined in order to decide nodes have to be pruned . If the weak connections are existed, the link connections are removed, and fuzzy rules are acquired from the network since the data procedure of the TAM network is similar to fuzzy logic. We discuss here the formulation of the pruning algorithm and show its usefulness through some examples.

2 TAM Network

The structure of TAM Network is shown in Figure 1. When feature maps, f_i, are given, the basis nodes are activated by the match between the activity distribution in a feature map, f_{ih}, and the distribution of the node's weights, w_{jih} and outputs to output nodes, y_j, are calculated as follows:

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

$$y_j = \prod_{i=1}^{M} x_{ji}. \quad (2)$$

where, ρ represents vigilance parameter and b_{ji} are inhibitory weights.

The class prediction, K, is the index of the maximally activated output node:

$$z_k = \sum_{j=1}^{N} y_j p_{jk} \quad (3)$$

$$K = \{k | \max_k z_k\}. \quad (4)$$

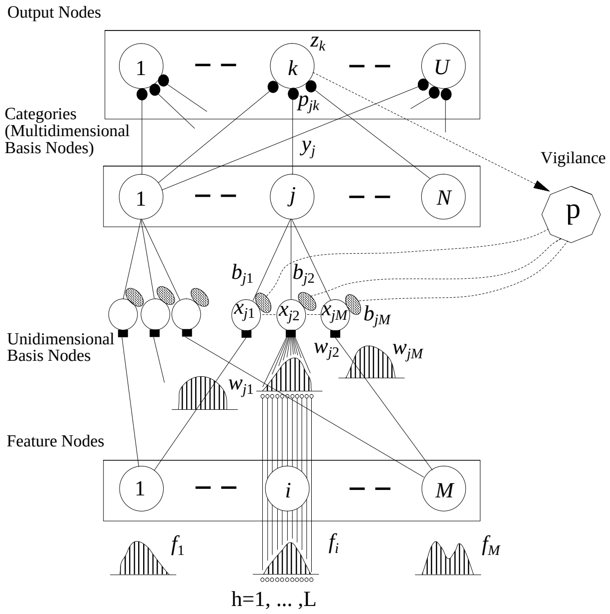

Figure 1: TAM Network

where, p_{jk} are weighted connections which represent the probability of each output given the category.

Let K^* denote the index of the "correct" supervised output class. If the network's output prediction K is not similar enough to the supervised output K^*, the "attention" is invoked and the ρ is raised until either the following subject is satisfied or until the maximal vigilance level, OC, is reached. When the ρ reached the maximal level, one node is added to categories.

$$\begin{array}{l} \text{If } \; z_{K^*}/z_K < OC \; \text{ then repeat} \\ \quad (a) \;\; \rho = \rho + \rho^{(step)} \\ \quad (b) \;\; \text{equation } (1)-(4) \\ \text{until either } \; z_{K^*}/z_K \geq OC \; \text{ or } \; \rho \geq \rho^{(max)} \end{array} \tag{5}$$

Once the subject of $z_{K^*}/z_K \geq OC$ is satisfied, the feedback signal y_j^* is calculated and the process moves to learning step.

$$z_k^* = 1 \text{ if } k = K^*; \quad z_k^* = 0 \text{ otherwise} \tag{6}$$

$$y_j^* = \frac{\prod_{i=1}^{M} x_{ji} \times \sum_{k=1}^{U} z_k^* p_{jk}}{\sum_{j'=1}^{N} \prod_{i=1}^{M} x_{j'i} \times \sum_{k=1}^{U} z_k^* p_{j'k}}. \tag{7}$$

The learning parameters, w_{jih}, p_{jk}, b_{ji}, are obtained as follows:

$$\triangle n_j = \alpha y_j^*(1-n_j) \tag{8}$$

$$\beta(M) = \frac{\lambda^{1/M}}{1-\lambda^{1/M}}, \qquad \lambda \in (0,1) \tag{9}$$

$$p_j^{(rate)} = \frac{\alpha}{\alpha + n_j} \tag{10}$$

$$\triangle w_{jih} = \frac{\alpha y_j^*(f_{ih} - w_{jih})}{\alpha\beta(M) + n_j} \tag{11}$$

$$\triangle p_{jk} = \frac{\alpha p_j^{(rate)} y_j^*(z_k^* - p_{jk})}{\alpha + n_j} \tag{12}$$

$$\triangle b_{ji} = b_j^{(rate)} y_j^*(x_{ji} - b_{ji}) \tag{13}$$

where, α, λ and $b_j^{(rate)}$ are parameters.

Whenever one datum put into the TAM network, the learning step is invoked and parameters are adjusted. The learning process is terminated when some epochs are achieved.

3 Pruning Algorithm

When the data set, D, in which we have the data, f_{si}, $s = 1, 2, \cdots, R$ with the class k, are given, all data, R, are divided into two group, i.e., learning data for learning the parameters and checking data for estimating the result.

The information entropy, $H(i)$, is calculated for estimating the significance of feature maps, and all of features are lined in order of value by $H(i)$.

$$H(i) = -\sum_{j=1}^{N} g_j \sum_{k=1}^{U} G_{jk} \log_2 G_{jk} \tag{14}$$

$$g_j = \frac{\sum_{s=1}^{R} x_{jis}}{\sum_{j=1}^{N} \sum_{s=1}^{R} x_{jis}} \tag{15}$$

$$G_{jk} = \frac{\sum_{s \in \psi_k} \gamma_{js} \times p_{jk}}{\sum_{s=1}^{R} \gamma_{js} \times p_{jk}} \tag{16}$$

$$\gamma_{js} = \prod_{i \in I^*} x_{jis} \times x_{jis} \tag{17}$$

where, ψ_k is a set of the data with the class k and I^* is a set of features chosen by $H(i)$.

The following feature i^* is extracted.

$$i^* = \{i \mid \max_i H(i)\} \tag{18}$$

In order to estimate the strength of link connections, the following three ruls are defined.

[Pruning Rule 1]

If the following condition is satisfied for checking data at each category j, the link connections between the category j and classes, k', $k' = 1, 2, \cdots, U$, $k' \neq k$, are removed. Simultaneously, the connections between the category j and features, $i' \notin I^*$, are removed.

$$G_{jk} \geq \eta \tag{19}$$

where, η is a threshold. We should notice the following features.

[Feature 1]
Even if the feature, i'', which satisfies $\sum_{s\in\psi_k} x_{ji''s} \geq \sum_{s\notin\psi_k} x_{ji''s}$ is selected at the next step, the equation (19) is also satisfied at that step.

[Pruning Rule 2]
If the following condition is satisfied for checking data at each category j, the link connections between the category j and features, i and $i' \notin I^*$, are removed.

$$\frac{1}{R}\sum_{s=1}^{R}\gamma_{js} < \theta \tag{20}$$

where, θ is a threshold.

[Feature 2]
After a category j is satisfied the equation (20) at once, the condition at the category is satisfied at the next step.

Now, we define the following φ_{jK} which means the cumulative degree of activation value of the data with the $K = K^*$ class at the j^{th} category in all of categories.

$$\varphi_{jK} = \frac{\sum_{s\in\Gamma_K}\gamma_{js}\times p_{jK}}{\sum_{j=1}^{N}\sum_{s\in\Gamma_K}\gamma_{js}\times p_{jK}} \tag{21}$$

where, $\Gamma_K = \{s|\ K = K^*,\ K = \max_k \sum_{j=1}^{N} y_j p_{jk}\}$.
We can line all of categories in order of their importance by φ_{jk} at the class k.

[Pruning Rule 3]
If the following condition is satisfied for checking data at the class K, the link connections between the class K and categories, $j',\ j' = 1, 2, \cdots, N,\ j' \neq j$, are removed.

$$\varphi_{jK} \geq \xi \tag{22}$$

where, ξ is a threshold.

[Feature 3]
In order to the maintain correctness at output nodes, we recommend to set the threshold ξ to the following value.

$$\xi \geq \max_{k\neq K}\frac{\sum_{s\in\Gamma_K}\sum_{j=1}^{N} p_{jk}}{\sum_{s\in\Gamma_K}\sum_{j=1}^{N} y_{js}p_{jk}} \tag{23}$$

The pruning of nodes and link connections is achieved according to the following algorithm.

[Step 1] For any feature i, calculate the following entropy $H(i)$ using checking data.

$$H(i) = -\sum_{j=1}^{N} g_j \sum_{k=1}^{U} G_{jk}\log_2 G_{jk} \tag{24}$$

$$g_j = \frac{\sum_{s=1}^{R} x_{jis}}{\sum_{j=1}^{N}\sum_{s=1}^{R} x_{jis}} \tag{25}$$

$$G_{jk} = \frac{\sum_{s\in\psi_k}\gamma_{js}\times p_{jk}}{\sum_{s=1}^{R}\gamma_{js}\times p_{jk}} \tag{26}$$

$$\gamma_{js} = \prod_{i\in I^*} x_{jis}\times x_{jis} \tag{27}$$

where, I^* is a set of i^* in the Step 2.

[Step 2] Select the most significant feature, i^*, which has

$$i^* = \{i|\max_i H(i)\}. \tag{28}$$

[step 3] If the following condition is satisfied for checking data by the first pruning rule, the link connections between the category j and classes $k' \neq k$ are removed. The link connections between the category j and features $i' \notin I^*$ are also removed.

$$G_{jk} \geq \eta \tag{29}$$

[Step 4] If the following condition is satisfied for checking data by the second pruning rule, the link connections between the category j and features i and $i' \notin I^*$ are removed.

$$\frac{1}{R}\sum_{s=1}^{R}\gamma_{js} < \theta \tag{30}$$

[Step 5] If the following condition is satisfied for checking data by the third pruning rule, the link connections between the class k and categories $j' \neq j$ are removed.

$$\varphi_{jk} \geq \xi \tag{31}$$

[Step 6] A category has lost connections to all classes or features is removed. Any class and feature has been disconnected from all categories are also removes.

[Step 7] Until all features are selected at step 2, let the algorithm repeat from step 1 to step 6.

When the algorithm is terminated, the optimum neural network pruned needless connections and nodes is obtained. We should notice that the algorithm is a kind of fuzzy tuning methods which adjust the number of features, classes and fuzzy rules since the data procedure is the same to the fuzzy logic.

4 Examples

4.1 Pattern Problem

We discuss here the availability of the pruning mechanism of the TAM network when the learning data in Figure 2 and Figure 3 are given.

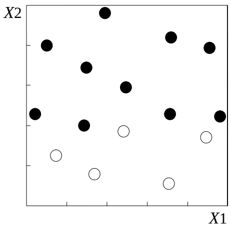

Figure 2: Learning Data 1

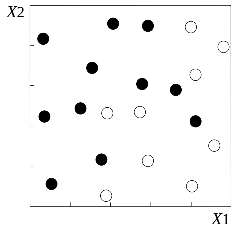

Figure 3: Learning Data 2

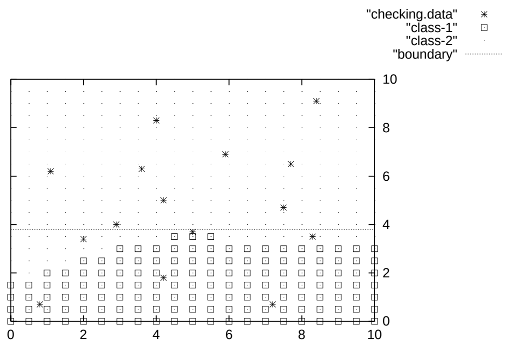

Figure 4: TAM Network

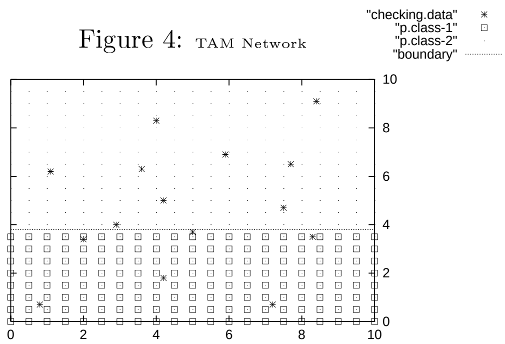

Figure 5: TAM Network after Pruning

The parameters for learning are set as follows:

$$
\begin{array}{lcllcl}
epoch & = & 10 & category\ init & = & 0 \\
L & = & 10 & rho\ init & = & 0.0 \\
OC & = & 0.8 & rho\ step & = & 0.1 \\
\alpha & = & 0.0000001 & rho\ max & = & 100.0 \\
\lambda & = & 0.33 & b_j^{(rate)} & = & 0.01.
\end{array}
$$

The correctness for the learning data of Figure 2 is obtained as 97.8% and for the learning data of Figure 3 as 81.7%. The correctness of the TAM network after pruning using checking data is compared with the TAM network before pruning. Tables 1 and 2 summarize the performance of the TAM network after pruning three times, which provide correctness and the number of nodes (in the feature-category-output layers). The correctness and the numbers of node in the TAM network before pruning are respectively 82.2% and 2-4-2 for Figure 2, and be 71.7% and 2-10-2 for Figure 3.

Table 1: Results of Checking Data 1

η	θ \ ξ	0.1		0.4		0.8	
0.005	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
	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
0.2	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
	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
0.5	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
	0.005	73.3	1-2-2	97.8	1-2-2	0.76	1-2-2
	0.0005	91.1	1-3.7-2	84.4	1-3.3-2	86.7	1-4-2

Table 2: Results of Checking Data 2

η	θ \ ξ	0.1		0.4		0.8	
0.005	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
	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
0.2	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
	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
0.5	0.05	0.0	0-0-2	0.0	0-0-2	0.0	0-0-2
	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

Figure 4 shows the checking data and an output of the TAM network for Figure 2. Figure 5 also shows the same checking data and an output of TAM network after pruning as $\eta = 0.005$, $\theta = 0.005$, and $\xi = 0.4$. Figure 6 shows the correct rate of the learning data and the checking data for the TAM network after pruning. The TAM network structure is shown in Figure 7. In the Figure 7, three fuzzy rules are extracted and the feature f_1 is removed.

The following three fuzzy rules are described by the TAM networks.

$$
\left.
\begin{array}{lllll}
r_1: & \text{if} & f_2 \text{ is } w_{32} & \text{then } K_2 = p_{32} & (0.624) \\
r_2: & \text{if} & f_2 \text{ is } w_{42} & \text{then } K_1 = p_{41} & (0.803) \\
r_3: & \text{if} & f_2 \text{ is } w_{42} & \text{then } K_2 = p_{42} & (0.197)
\end{array}
\right\} \quad (32)
$$

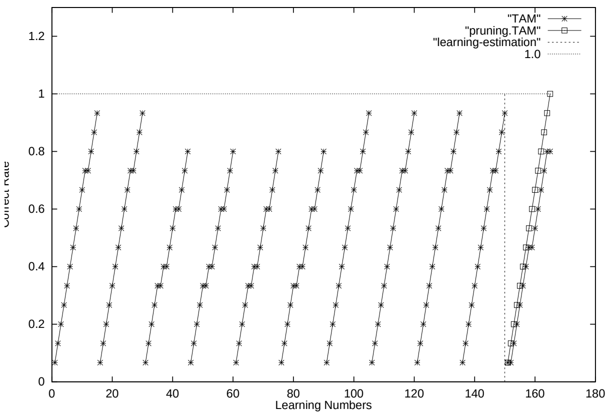

Figure 6: Correct Rate

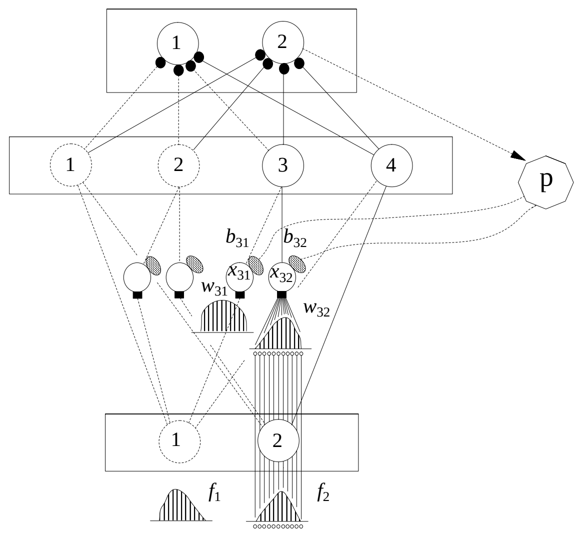

Figure 7: TAM Network after Pruning

4.2 Aperture Problem

We discuss here the necessity of feedback mechanism of the TAM network. The aperture problem is a kind of the visual experiment, where one circle aperture is first displayed and a linear line is next moving to a direction inside the aperture. While the line of the first circle is moving, other two circles appear at both sides of the first circle appear, and the lines in the second circles are moving for the different direction from the first line. After both of the circles appeared, the direction of the first line is seemed to be changed itself for the same direction of lines of the both sides(Figure 8).

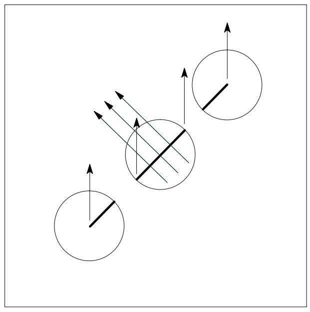

Figure 8: Aperture Experiment

Okada[12] explained by an experiment that the perception might to be tending upward according to speed slowly of moving lines. He also discussed the necessity of feedback mechanism in the visual system.

Figure 9 shows that the perception is tending upward according to the displaying time. Four cases, $(35, 90)$, $(33, 86)$, $(38, 96)$ and $(40, 100)$, where (circle radius, distance between centers), are experimented. The subjects of the experiment react as same as the case of Okada.

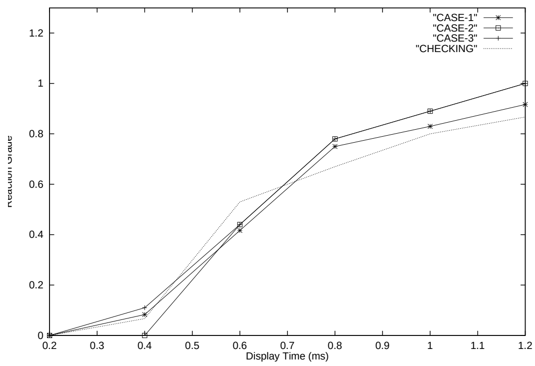

Figure 9: Reaction of Subjects

The reaction of subjects is expected to depend on strongly the gap scale between apertures. The gap scale of experiments in Figure 9 was 20mm, and therefore the four perceptions might to be the same. In order to confirm the necessity of feedback mechanism of the TAM network, let the three cases, $(35, 90)$, $(33, 86)$ and $(38, 96)$, to be the learning data for TAM network, and (40,100) to be the checking data. The parameters for learning are set as follows:

$$
\begin{array}{lllllll}
epoch & = & 3 & \quad & category\ init & = & 0 \\
L & = & 10 & & rho\ init & = & 0.0 \\
OC & = & 0.8 & & rho\ step & = & 0.1 \\
\alpha & = & 0.0000001 & & rho\ max & = & 100.0 \\
\lambda & = & 0.33 & & b_j^{(rate)} & = & 0.01
\end{array}
$$

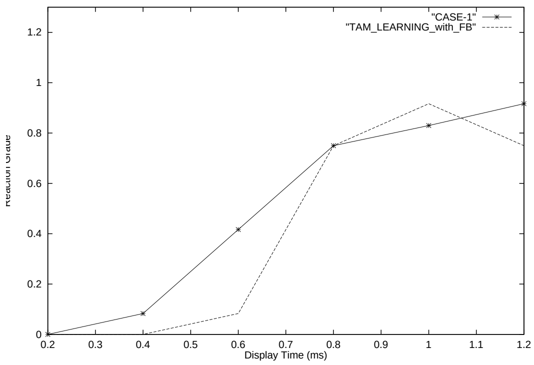

Figure 10: Reaction after Learning by TAM Network

Figure 10 shows the output of TAM in the case of (35,90). Table 3 shows the correct rate of the learning data (35,90) and the checking data (40,100) as $\eta = 0.5$, $\theta = 0.005$, and $\xi = 0.1$ for pruning parameters. To cut off the feedback mechanism in the TAM network, we let the vigilance learning be inactive. The correctness of the TAM network after pruning is slightly better than before pruning. The correctness without feedback mechanism is worse for checking data than with feedback mechanism. The 104 fuzzy rules are extracted from the TAM network with feedback mechanism after pruning.

Table 3: Results of TAM Network

		TAM	Pruning TAM
with FB	Learning Data	0.8278	0.85
	Checking Data	0.578	0.65
without FB	Learning Data	0.861	0.867
	Checking Data	0.544	0.6

5 Conclusions

We proposed here the pruning mechanism of the TAM network for acquiring fuzzy rules, and discussed its usefulness and the necessity of feedback mechanism.

References

[1] M.Matsumura, "Introduction to Neuroscience", Science Publishing Co. (1995) in Japanese

[2] S.Grossberg and E.Mingolla: "Neural Dynamics of Perceptual Grouping: Texture, Boundaries, and Emergent Segmentation," *Percept Psychophys*, Vol.38, pp.141-171 (1985)

[3] 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)

[4] H.Neumann and W.Sepp: "Recurrent V1-V2 Interaction in Early Visual Boundary Processing," *Biological Cybernetics*, Vol.81, pp.425-444 (1999)

[5] 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)

[6] G.A.Carpenter, S.Grossberg, N.Markuzon, J.Reynolds and D.B.Rosen: "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Learning of Analog Multidimensional Maps," *IEEE Transactions on Neural Networks*, Vol.3, pp.698-713 (1992)

[7] J.R.Williamson, "Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps," *Neural Networks*, Vol.9, pp.881-897 (1996)

[8] J.R.Williamson: "Self-organization of Topographic Mixture Networks Using Attentional Feedback," *Technical Report CAS/CNS TR-99-027* (1999)

[9] R.Reed: "Pruning Algorithms - A Survey," *IEEE Transactions on Neural Networks*, Vol.4, pp.740-747 (1993)

[10] 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)

[11] I.Hayashi, "Acquisition of Fuzzy Rules Using Fuzzy ID3 with Ability of Learnining for AND/OR Operators," *1996 Australian New Zealand Conference on Intelligent Information Systems*, pp.187-190 (1996)

[12] K.Fukushima, Y.Yamaguchi and M.Okada, "Neural Network Model of the Spatial Memory: Associative Recall Maps," it Neural Networks, Vol.10, pp.971-979 (1997)