A Formulation of Knowledge Restructuring Type TAM Network *

Isao Hayashi
Graduate School of Corporate Information
Hannan University
Matsubara City, Osaka, Japan
ihaya@kcn.hannan-u.ac.jp

James R. Williamson
Theater Missile Defense, Group 32, MIT
Lockheed Martin Corporation
Lexington, MA, U.S.A.
jrw@ll.mit.edu

Abstract – *The TAM (Topographic Attentive Mapping) network is a biologically-motivated neural network. In the TAM network algorithm, knowledge is acquired from the TAM network structure through a pruning method, and the acquired knowledge is stored in an independent module. In this paper, three measures of similarity used to compare these independent knowledge storage modules will be discussed.*

Keywords: Neural Network, Knowledge, Fuzzy Sets, Restructuring.

1 Introduction

Several models which translate the neural network structure into knowledge representation have been proposed [8], [9], [5]. However, these models only rehash the neural network structure as knowledge representaion. There isn't enough discussion how to built groupings of knowledge and how to restructure the knowledge representation. Fischer[3] claims knowledge management means a loop process of creation, integration and dissemination. Nonaka[11] claims there are four stages of knowledge representation which are Socialization, Externalization, Combination and Internalization. Amitani[1] has explained a method of knowledge creation and reconstruction. Carpenter[2] has mentioned that even humans cannot easily correlate knowledge representations (class) of instances acquired by experience with a class name when he / she first experience that instance. For example, children can distinguish and gather similar shapes of drink holders, e.g., cup, glass and mug, but they can't tell whether these holder should be called as "cups", "glasses" or something else until their mother teaches them what the class names are (See Figure 1). In general, a class name is given while denoting and item to the represent a grouping or a concept after instances are collected and grouped. Thus, the important thing is how to group the instances and restructure the instances, not name the class.

In this paper, we propose a new knowledge system with restructuring mechanism of knowledge. We called the knowledge system the knowledge restructuring type TAM network (shortly, KRT-TAM network). Figure 2 shows the diagram of the KRT-TAM network. TAM (Topographic Attentive Mapping) network[12], [6], [7] is based on a biologically-motivated framework and possesses a powerful learning mechanism[4], [10]. TAM's feature layer is constructed to imitate the retina, the category layer imitates the LGN (lateral geniculate nucleus) and the output layer imitates visual cortex. After learning the TAM network, knowledge representation acquired from the TAM network are memorized to a knowledge part as a formula of fuzzy rules. The knowledge part is constructed by classes, cases and fuzzy rules (instances). The fuzzy rules are restructured whenever the incremental learning of the TAM network is accomplished. We discuss here how to formulate the KRT-TAM network and show the usefuleness of the KRT-TAM network through concrete examples.

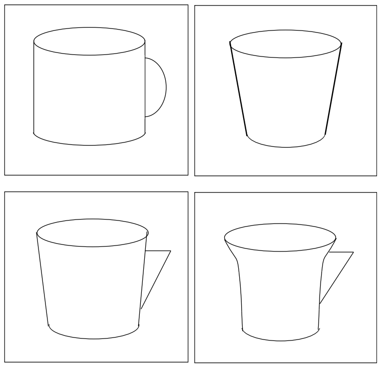

Figure 1: Examples of Cups and Glasses

2 TAM Network

The structure of the TAM Network is shown in Figure 3. When feature maps, f_{ih}, are given, the output signal to the category layer, y_j, are calculated using the node's weights, w_{jih}.

$$\begin{aligned} y_j &= \prod_{i=1}^{M} x_{ji} \\ &= \prod_{i=1}^{M} \frac{\sum_{h=1}^{L} f_{ih} w_{jih}}{1+\rho^2 b_{ji}} \end{aligned} \quad (1)$$

*0-7803-7952-7/03/$17.00 ©2003 IEEE.

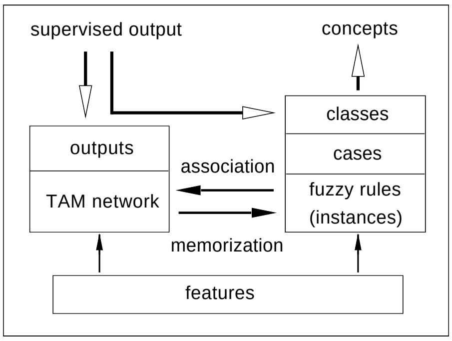

Figure 2: Concept of KRT-TAM Network

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

The output prediction, K, is calculated as follows:

$$\begin{aligned} K &= \{k| \max_k z_k\} \\ &= \{k| \max_k \sum_{j=1}^{N} y_j p_{jk}\} \end{aligned} \tag{2}$$

where, z_k are the output at each node of output layer and p_{jk} are weighted connections.

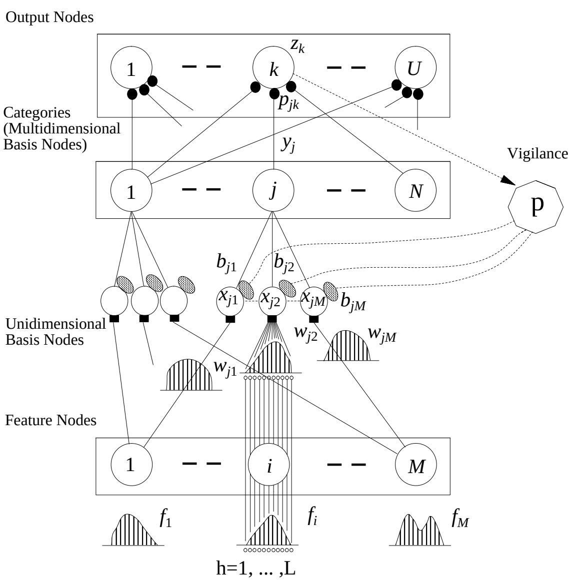

Figure 3: TAM Network

Let K^* denote the index of the "correct" supervised output class. If the network's output prediction K is not similar enough to K^*, we do $\rho = \rho + \rho^{(step)}$ until either $z_{K^*}/z_K \geq OC$ or $\rho \geq \rho^{(max)}$, where OC is the maximal vigilance level. Once the subject of $z_{K^*}/z_K \geq OC$ is satisfied, the feedback signal y_j^* is calculated for the learning step.

$$y_j^* = \frac{\prod_{i=1}^{M} x_{ji} \times \sum_{k=1}^{U} z_k^* p_{jk}}{\sum_{j'=1}^{N} \prod_{i=1}^{M} x_{j'i} \times \sum_{k=1}^{U} z_k^* p_{j'k}} \tag{3}$$

$$z_k^* = 1 \text{ if } k = K^*; \quad z_k^* = 0 \text{ otherwise} \tag{4}$$

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

$$\triangle w_{jih} = \frac{\alpha y_j^* (1 - \lambda^{1/M})(f_{ih} - w_{jih})}{(\alpha - 1)\lambda^{1/M} + n_j}, \lambda \in (0,1) \tag{5}$$

$$\triangle p_{jk} = \frac{\alpha y_j^* (z_k^* - p_{jk})}{\alpha + n_j} \tag{6}$$

$$\triangle b_{ji} = b_j^{(rate)} y_j^* (x_{ji} - b_{ji}) \tag{7}$$

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

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

The algorithm of the TAM network including learning steps and pruning steps is represented as follows:

[Step 1] The output prediction, K, is calculated.
[Step 2] If K is not similar enough to K^*, we do $\rho = \rho + \rho^{(step)}$. When ρ reachs the maximal level, one node is added to categories.
[Step 3] If $z_{K^*}/z_K \geq OC$, the learning step starts. Parameters, w_{jih}, p_{jk} and b_{ji}, are updated.
[Step 4] Until $z_{K^*}/z_K \geq OC$, let the algorithm repeat from step 1 to step 3.
[Step 5] After learning, the pruning step starts. The data set in which $f_{si}, \ s = 1, 2, \cdots, R$ is divided into learning data and checking data. The information entropy, $H(i)$, is calculated using the learning data for feature selections, where ψ_k is a set of the data of the class k.

$$H(i) = -\sum_{j=1}^{N} g_j \sum_{k=1}^{U} G_{jk} \log_2 G_{jk} \tag{9}$$

$$g_j = \frac{\sum_{s=1}^{R} x_{jis}}{\sum_{j=1}^{N} \sum_{s=1}^{R} x_{jis}} \tag{10}$$

$$G_{jk} = \frac{\sum_{s \in \psi_k} \gamma_{js} \times p_{jk}}{\sum_{s=1}^{R} \gamma_{js} \times p_{jk}} \tag{11}$$

$$\gamma_{js} = \prod_{i \in I^*} x_{jis} \times x_{jis} \tag{12}$$

[Step 6] The following feature i^* is extracted as an important feature and we set $I^* = \{i^*\}$.

$$i^* = \{i| \max_i H(i)\} \tag{13}$$

[Step 7] If the following condition is satisfied for checking data at a category j, the link connections between j and outputs $k', \ k' = 1, 2, \cdots, U, \ k' \neq k$, are removed. Simultaneously, the connections between j and features $i' \notin I^*$, are removed, where η is a threshold.

$$G_{jk} \geq \eta \tag{14}$$

[Step 8] If the following condition is satisfied for checking data at the category j, the link connections between j and i, and $i' \notin I^*$, are removed, where θ is a threshold.

$$\frac{1}{R}\sum_{s=1}^{R}\gamma_{js} < \theta \tag{15}$$

[Step 9] If the following condition is satisfied for checking data at K, the link connections between K and categories, $j', \ j' = 1, 2, \cdots, N, \ j' \neq j$, are removed, where ξ is a threshold.

$$\varphi_{jK} = \frac{\sum_{s\in\Gamma_K}\gamma_{js}\times p_{jK}}{\sum_{j=1}^{N}\sum_{s\in\Gamma_K}\gamma_{js}\times p_{jK}} \geq \xi \tag{16}$$

[Step 10] When a category has lost connections to all outputs or features, the category is removed. Any output and feature which has been disconnected from all categories is also removed.
[Step 11] Until all features are selected at step 6, let the algorithm repeat from step 5 to step 10.

When the algorithm is terminated, the neural network whose needless connections and nodes are pruned is obtained. We should notice that the algorithm is a kind of fuzzy tuning methods since the data procedure is the same as that of fuzzy logic. Thus, we can acquire fuzzy rules from the TAM network as a knowledge representaion.

3 Knowledge of the KRT-TAM Network

The knowledge acquired from the TAM network are memorized to knowledge part as the following fuzzy rule format:

$$\left.\begin{array}{ll} r_1: & \text{if } f_1 \text{ is } z_{11} \text{ and } \cdots \text{ and } f_M \text{ is } z_{1M} \\ & \text{then } C_1 = o_{11}, \ \cdots, \ C_U = o_{1U} \\ r_2: & \text{if } f_1 \text{ is } z_{21} \text{ and } \cdots \text{ and } f_M \text{ is } F_{2M} \\ & \text{then } C_1 = o_{21}, \ \cdots, \ C_U = o_{2U} \\ & \vdots \\ r_N: & \text{if } f_1 \text{ is } z_{N1} \text{ and } \cdots \text{ and } f_M \text{ is } z_{NM} \\ & \text{then } C_1 = o_{N1}, \ \cdots, \ C_U = o_{NU} \end{array}\right\} \tag{17}$$

where z_{ji} $j = 1, 2, \cdots, N, \ i = 1, 2, \cdots, M$ are the fuzzy sets, which were the node's weights, w_{ji}, in TAM network. The resultant numbers, o_{jk}, were the weighted connections of output nodes, p_{jk}, of the TAM network.

The each rule is ranked according to φ_{jk} of the equation (16) as follows:

$$\left.\begin{array}{ll} C_1: & r_1(\varphi_{11}) \geq r_2(\varphi_{21}) \geq \cdots \geq r_N(\varphi_{N1}) \\ C_2: & r_1(\varphi_{12}) \geq r_2(\varphi_{22}) \geq \cdots \geq r_N(\varphi_{N2}) \\ & \vdots \\ C_U: & r_1(\varphi_{1U}) \geq r_2(\varphi_{2U}) \geq \cdots \geq r_N(\varphi_{NU}). \end{array}\right\} \tag{18}$$

In the knowledge part, we have "Class" and "Case". A case means a general structure of knowledge acquired from the TAM network. The similar cases are grouped and the superviser then gives the group a class name. Figure 4 shows an example, in which there are two classes, "Cups" with three cases and "Glasses" with two cases.

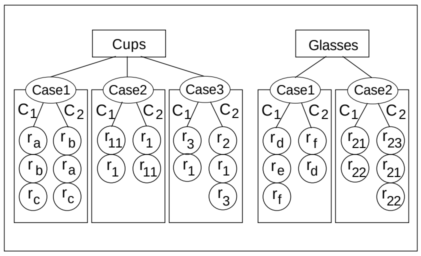

Figure 4: Example of Knowledge Part

When a knowledge representaion of fuzzy rules moves to the knowledge part from the TAM network, we need to define a similarity measure between the knowledge and cases which exist in the knowledge part in order to decide which class should incorporate the knowledge. We define here the following three types of similarity measures.

$$\begin{aligned} 1)\ \Omega &= \begin{cases} -: & -\tau < I_1^* < 0 \\ +: & 0 \leq I_1^* < \tau \\ F: & |I_1^*| \geq \tau \end{cases} \qquad (19) \\ I_1^* &= \{I_1 \mid min_{Case}|I_1|\} \\ I_1 &= \frac{1}{UNL}\sum_{k=1}^{U}\sum_{j=1}^{N}\sum_{i=1}^{M+1}\sum_{h=1}^{L}\log\frac{\mu_{w_{ji}}(f_{ih})}{\mu_{z_{ji}}(f_{ih})} \end{aligned}$$

$$\begin{aligned} 2)\ \Omega &= \begin{cases} -: & max_{\{-,+,F\}}\ count_{k,j}\ \omega = \{-\} \\ +: & max_{\{-,+,F\}}\ count_{k,j}\ \omega = \{+\} \\ F: & max_{\{-,+,F\}}\ count_{k,j}\ \omega = \{F\} \end{cases} \qquad (20) \\ \omega &= \begin{cases} -: & -\tau < I_2^* < 0 \\ +: & 0 < I_2^* < \tau \\ F: & |I_2^*| \geq \tau \end{cases} \\ I_2^* &= \{I_2 \mid min_{Case}|I_2|\} \\ I_2 &= \frac{1}{L}\sum_{i=1}^{M+1}\sum_{h=1}^{L}\log\frac{\mu_{w_{ji}}(f_{ih})}{\mu_{z_{ji}}(f_{ih})} \end{aligned}$$

$$\begin{aligned} 3)\ \Omega &= \begin{cases} -: & max_{\{-,+,F\}}\ count_{k,j\in N'}\ \omega = \{-\} \\ +: & max_{\{-,+,F\}}\ count_{k,j\in N'}\ \omega = \{+\} \\ F: & max_{\{-,+,F\}}\ count_{k,j\in N'}\ \omega = \{F\} \end{cases} \qquad (21) \\ \omega &= \begin{cases} -: & -\tau < I_3^* < 0 \\ +: & 0 < I_3^* < \tau \\ F: & |I_3^*| \geq \tau \end{cases} \\ I_3^* &= \{I_3 \mid min_{Case}|I_3|\} \\ I_3 &= \frac{1}{L}\sum_{i=1}^{M+1}\sum_{h=1}^{L}\log\frac{\mu_{w_{ji}}(f_{ih})}{\mu_{z_{ji}}(f_{ih})} \end{aligned}$$

where, $\mu_{w_{jM+1}}(f_{M+1}) = p_{jk}$ and $\mu_{z_{jM+1}}(f_{M+1}) = o_{jk}$, the "$count_{k,j}\ \omega$" means the number of counting ω in the class k and j, and $N' \leq N$ and τ is a threshold.

The algorithm for resturucturing cases is achieved according to the following steps.

[Step 1] Rank the knowledge acquired from the TAM network according to φ_{jk} of the equation (18).
[Step 2] Calculate the similarity measure Ω among the equations (19) to (21).
[Step 3] In the case of $\Omega = \{F\}$, the acquired knowledge is registered as a case of a new class. The learning of the TAM network is continued.
[Step 4] In the case of $\Omega = \{+\}$, the acquired knowledge is registered as a new case of the existing class of $I_g^*,\ g = 1, 2, 3$. The learning of the TAM network is continued.
[Step 5] In the case of $\Omega = \{-\}$, the acquired knowledge is presumed to be the same as the case of I_g^*. We let the TAM network relearn from b_{ji}, $w_{ji}(z_{ji})$ and $p_{jk}(o_{jk})$ of the case of I_g^* as the initial parameter values. After learning, the resultant knowledge of the TAM network is exchanged to the existing case in the knowledge part.

4 Examples

In this section, three types of similarity measures are compared each other through an image example and an example of how to restructure cases is presented.

The ellipse image shown in (a) of Figure 6 is compared with two kinds of original images, circle and square, shown in (a) and (b) of Figure 5. The comparison is using knowledge representation of fuzzy rules acquired from TAM network. The output images of TAM network after learning are shown in (c) and (d) of Figure 5, and in (b) of Figure 6, respectively. The parameters of the TAM network are set as follows:

L	=	10	$rho\ init$	=	0.0
OC	=	0.8	$rho\ step$	=	0.1
α	=	0.0000001	$rho\ max$	=	100.0
λ	=	0.33	$b_j^{(rate)}$	=	0.01
η	=	0.8	θ	=	0.03
ξ	=	0.5	N_{set}	=	6.

The result is shown in Table 1, Table 2 and Table 3. The eight fuzzy rules are acquired from the circle image, eleven from the square image and six from the ellipse image. All similarity measure indicate that the ellipse image is more similar to circle than to square. However, both the second similarity measure and the third similarity measure depend on the threshold of τ value, so we should discuss the problems more in the near future.

Next, we show an example of how to restructure cases using here the first similarity measure. Three data sets which include "△" and "•" are shown in Figure 7. The boundaries between "△" and "•" in the first data set and the third data set are represented by around the funcion $y = x$. Alternatively, the boundary of the second data set is $y = -x$. The

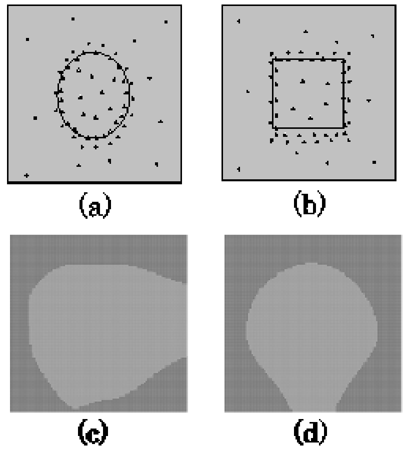

Figure 5: Circle Image and Square Image

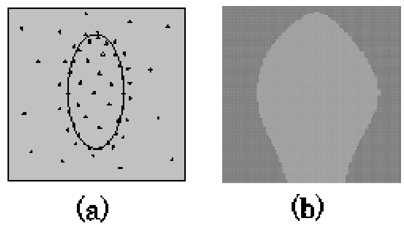

Figure 6: Ellipse Image

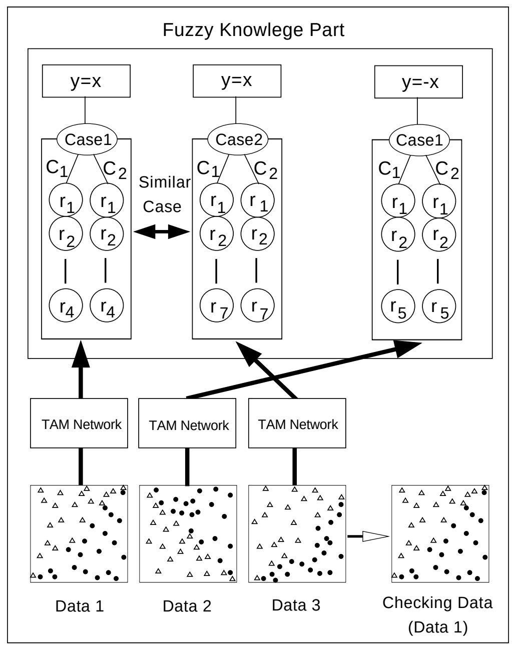

Figure 7: Procedure of the Fuzzy TAM Network

Table 1: Result of Similarity Measure 1

	Similarity Measure 1
Similarity between Circle and Ellipse	32.45
Similarity between Square and Ellipse	65.05

Table 2: Result of Similarity Measure 2

	Similarity Measure 2			
	+	-	F	τ
Similarity between Circle and Ellipse	10	0	6	5.0
	12	0	4	7.5
Similarity between Square and Ellipse	0	0	24	5.0
	6	0	18	7.5

parameters of the TAM network are set as follows:

$$\begin{array}{lcllcl} L & = & 10 & rho\ init & = & 0.0 \\ OC & = & 0.8 & rho\ step & = & 0.1 \\ \alpha & = & 0.0000001 & rho\ max & = & 100.0 \\ \lambda & = & 0.33 & b_j^{(rate)} & = & 0.01 \\ \eta & = & 0.8 & \theta & = & 0.03 \\ \xi & = & 0.5 & \tau & = & 10.0. \end{array}$$

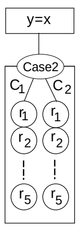

Figure 8: A New Case 1

First, the fuzzy rule for the first data set is acquired from the TAM network and the fuzzy rule is memorized as a case, "Case 1", of "Class($y = x$)" in the knowledge part. Next, the fuzzy rule for the second data set is acquired from the TAM network and moved into the knowledge part. The first similarity measure of the second fuzzy rule with "Case 1" is obtained as $I_1^*(2,1) = 13.5$ as shown in Table 4. Since τ is set to 10.0, $\omega = \{F\}$ and the fuzzy rule of the second data set is registered as "Case 1" of a new class, "Class($y = -x$)". When the third fuzzy rule for the third data set is moved into the knowledge part, the similarity measures of the third fuzzy rule with the first case and the second case are calculated, respectively. Since $I_1(3,1) = -2.84$, $I_1(3,2) = -1.88$ and $I_1^*(3,1) = -1.88$, $\omega = \{-\}$ and we let the TAM network relearn from b_{ji}, $w_{ji}(z_{ji})$ and $p_{jk}(o_{jk})$ of the first case as the initial parameter values. The resultant fuzzy rule TAM network is exchanged with the first case and registered as "Case 1". The new case, "Case 1", is shown in Figure 8.

Table 3: result of Similarity Measure 3

	Similarity Measure 3			
	+	-	F	τ
Similarity between Circle and Ellipse	10	0	2	5.0
	12	0	0	7.5
Similarity between Square and Ellipse	0	0	12	5.0
	6	0	6	7.5

Table 4: Result of I(W,Z)

	Case 1 in Knowledge Part	Case 2 in Knowledge Part
Data 2	13.5	–
Data 3	-2.84	-1.88

To show the usefulness of the KRT-TAM network by restructuring cases, the correct rate for checking data, which is the same as the first data, is compared to the conventional TAM network. The result is shown in Table 5. The correct rate of the conventional TAM network is obtained as 50.0% and the number of categories is 7. On the other hand, the correct rate of the KRT-TAM network is obtained as 60.0% and the number of categories is 5. The correct rate of the KRT-TAM network is better than the TAM network and the number of categories is less than TAM network.

Table 5: Correct Rate and Categories

	Correct Rate		Categories	
	TAM	KRT TAM	TAM	KRT TAM
Data 1	97.5%	97.5%	4	4
Data 2	80.0%	80.0%	5	5
Data 3	80.3%	–	7	–
Data 3 after relearning	–	97.5%	–	6
Checking Data	50.0%	60.0%	7	5

5 Conclusions

We formulated here KRT-TAM network and showed the usefulness of KRT-TAM network through examples.

This research is partially supported by the Ministry of Education, Culture, Sports, Science and Technology of

Japan under Grant-in-Aid for Scientific Research number 14580433.

References

[1] S.Amitani, M.Mori and K.Hori, "An approach to the supporting system for event planning - toward a knowledge reconstruction engine -", *I*CONIP'02-SEAL'02-FSKD'02, pp. 208, 2002.

[2] G.A.Carpenter and S.Grossberg and J.Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network", *N*eural Networks, Vol.4, pp. 565–588, 1991.

[3] G.Fischer and J.Ostwald, "Knowledge management: Problems, promises, realities, and challenges", *I*EEE Intelligent Systems, Vol.16, No.1, pp. 60–72, 2001.

[4] S.Grossberg, "How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex", *S*patial Vision, Vol.12, No.2, pp. 163–185, 1999.

[5] I.Hayashi and M.Umano and T.Maeda and A.Bastian and L.C.Jain, "Acquisition of fuzzy knowledge by NN and GA - A survey of the fusion and union methods proposed in Japan", Proc. Second International Conference on Knowledge-Based Electronic Systems, (KES'98), pp. 69–78, 1998.

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

[7] I.Hayashi and J.R.Williamson, "An analysis of aperture problem using fuzzy rules acquired from TAM networks", Proc. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2002), pp. 914–919, 2002.

[8] B.Kosko, *N*eural networks and fuzzy systems, A dynamical approach to machine intelligence, Prentice-Hall, 1992.

[9] K.Nakaoka and T.Furuhashi and Y.Uchikawa, "A study on apportionment of credits of fuzzy classifier system for knowledge acquisition of large scale systems", Proc. The third IEEE International Conference on Fuzzy Systems, Vol.2, pp. 1797–1800, 1994.

[10] H.Neumann and W.Sepp, "Recurrent V1-V2 interaction in early visual boundary processing", *B*iological Cybernetics, Vol.81, pp. 425–444, 1999.

[11] I.Nonaka and H.Takeuchi, *T*he knowledge-creating company: How Japanese companies create the dynamics of innovation, Oxford University Press, 1995.

[12] J.R.Williamson, "Self-organization of topographic mixture networks using attentional feedback", *N*eural Computation, Vol.13, pp. 563–593, 2001.