KNOWLEDGE RESTRUCTURING IN FUZZY TAM NETWORK

Isao Hayashi

Hannan University
5-4-33, Amami-higashi, Matsubara
Osaka 580-8502, Japan

James R. Williamson

Lockheed Martin Corp.
Theater Missile Defense, group 32, MIT
Lexington, MA 02420, U.S.A.

ABSTRACT

TAM network (Topographic Attentive Mapping network) is a biologically-motivated neural network. With the pruning algorithm, fuzzy rules are acquired from the TAM network sturucture. In this paper, the restructuring algorithm of fuzzy rules is discussed and the usefuleness of the algorithm is illustrated.

1. INTRODUCTION

Several neural networks based on the human visual system have been proposed, e.g., BCS[1], Neocognitron[2], ARTMAP[3], TAM[4], [5], Visual Cortex Model[6], V1-V2 model[7]. TAM (Topographic Attentive Mapping) network is an especially significant model since the structure is based on a biologically-motivated framework and possesses a powerful learning mechanism. 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. By applying the pruning algorithm to these layers, fuzzy rules are acquired from TAM network structure[5], [11].

On the other hand, several models which translate the neural network structure into knowledge representation have been proposed[8], [9], [10]. However, these models only rehash the neural network structure as knowledge representaion. There isn't enough discussion how to built groupings of knowledge representation and how to restructure the knowledge. Carpenter[3] 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). Thus, 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

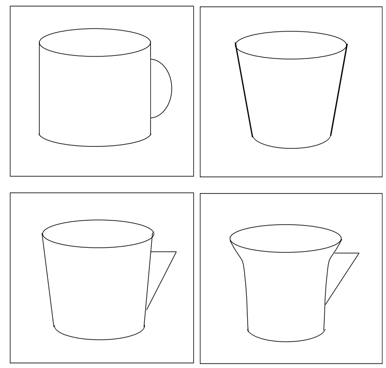

Figure 1: Examples of Cups and Glasses

the instances, not name the class. Moreover, humans might not have the correct concept for the class. In our model, we propose that the characteristic by which categorizes are divided are vague and are integrated into the concept in a gradual cumulative process (See Figure 2).

In this paper, we propose a new knowledge system including the TAM network and a restructuring mechanism of fuzzy rules. we called the system the fuzzy TAM network. Figure 3 shows the diagram of the fuzzy TAM network. After learning the TAM network, fuzzy rules acquired from the TAM network are memorized to a fuzzy knowledge part. The fuzzy 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 fuzzy TAM network and show the usefuleness of the fuzzy TAM network through a concrete example.

2. TAM NETWORK

The structure of the TAM Network is shown in Figure 4. When feature maps, f_{ih}, are given, the output signal to the category layer, y_j, are calculated using the node's weights,

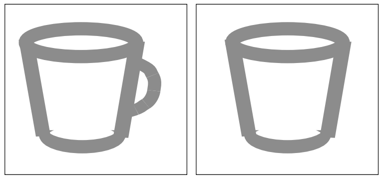

Figure 2: Concept of Cups and Glasses

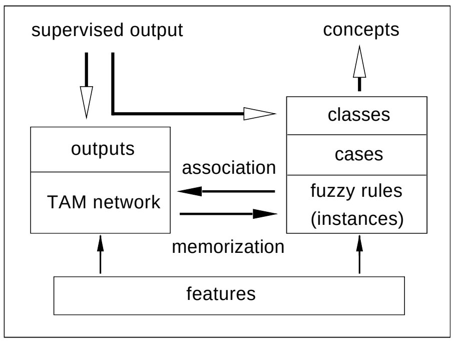

Figure 3: Concept of Fuzzy TAM Network

w_{jih}.

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

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

The output prediction, K, is calculated as follows:

$$K = \{k \mid \max_k z_k\} = \{k \mid \max_k \sum_{j=1}^{N} y_j p_{jk}\} \tag{2}$$

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

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.

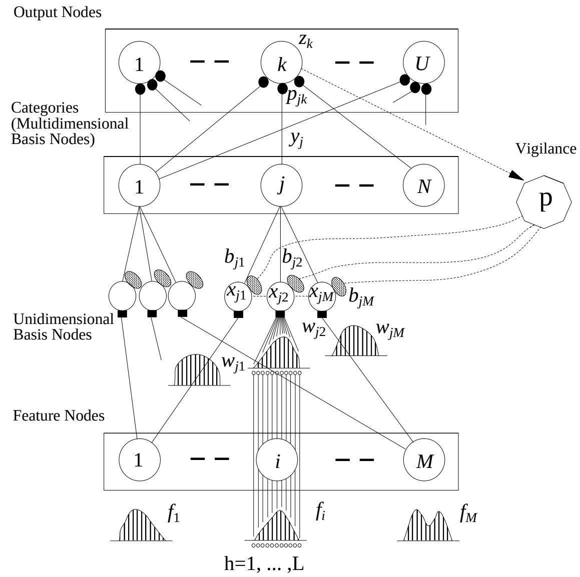

Figure 4: TAM Network

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

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

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 $I^* = \{i^*\}$.

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

[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.

3. FUZZY KNOWLEDGE OF THE TAM NETWORK

The fuzzy rules acquired from the TAM network are memorized to fuzzy knowledge part by the following 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 membership functions, which were the node's weights, w_{ji}, in TAM network. The real numbers, o_{jk}, were the weighted connections of output nodes, p_{jk}, of the TAM network.

The fuzzy rules are 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 fuzzy knowledge part, we have "Class" and "Case". A case means a general structure of fuzzy rules acquired from the TAM network. We call the case "fuzzy case". The similar cases are grouped and the superviser then gives the group a class name. Figure 5 shows an example, in which there are two classes, "Cups" with three fuzzy cases and "Glasses" with two fuzzy cases.

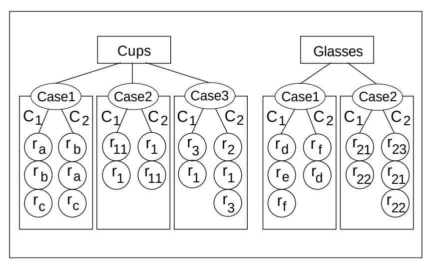

Figure 5: Example of Fuzzy Knowledge Part

Now, when a fuzzy rule moves to the fuzzy knowledge part from the TAM network, we need to define a similarity measure between the fuzzy rules and fuzzy cases which exist in the fuzzy knowledge part to decide which class should incorporate the fuzzy rule. We denote an M dimensional membership function of fuzzy rules as $\mu_{\mathbf{W_j}}(\mathbf{f})$ and an M

dimensional membership function of fuzzy case as $\mu_{\mathbf{Z_j}}(\mathbf{f})$.

$$\mu_{\mathbf{W_j}}(\mathbf{f}) = \mu_{w_{j1},w_{j2},\cdots,w_{jM}}(f_1, f_2, \cdots, f_M) \quad (19)$$
$$\mu_{\mathbf{Z_j}}(\mathbf{f}) = \mu_{z_{j1},z_{j2},\cdots,z_{jM}}(f_1, f_2, \cdots, f_M) \quad (20)$$

We define the similarity measure as follows:

$$\begin{aligned} I(W,Z) &= \frac{1}{UN}\sum_{k=1}^{U}\sum_{j=1}^{N} E\{\log \frac{\mu_{\mathbf{W_j}}(\mathbf{f}) \times p_{jk}}{\mu_{\mathbf{Z_j}}(\mathbf{f}) \times o_{jk}}\} \\ &= \frac{1}{UN}\sum_{k=1}^{U}\sum_{j=1}^{N}\sum_{i=1}^{M+1} E\{\log \frac{\mu_{w_{ji}}(f_i)}{\mu_{z_{ji}}(f_i)}\} \\ &= \frac{1}{UNL}\sum_{k=1}^{U}\sum_{j=1}^{N}\sum_{i=1}^{M+1}\sum_{h=1}^{L} \log \frac{\mu_{w_{ji}}(f_{ih})}{\mu_{z_{ji}}(f_{ih})} \end{aligned} \quad (21)$$

where, $\mu_{w_{jM+1}}(f_{M+1}) = p_{jk}$ and $\mu_{z_{jM+1}}(f_{M+1}) = o_{jk}$.

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

[Step 1] Rank the fuzzy rule acquired from the TAM network according to φ_{jk} of the equation (18).

[Step 2] Calculate the similarity measure $I(W,Z)$ of the equation (21).

[Step 3] Calculate the minimized similarity measure, $I^*(W,Z) = \{I(W,Z) \mid min|I(W,Z)|\}$, among all fuzzy cases.

[Step 4] In the case of $|I^*(W,Z)| \geq \tau$, where τ is a threshold, the fuzzy rule is registered as a fuzzy case of a new class. The learning of the TAM network is continued.

[Step 5] In the case of $0 \leq I^*(W,Z) < \tau$, the fuzzy rule is registered as a new fuzzy case of the existing class of $I^*(W,Z)$. The learning of the TAM network is continued.

[Step 6] In the case of $-\tau < I^*(W,Z) < 0$, the fuzzy rule is presumed to be the same as the fuzzy case of $I^*(W,Z)$. We let the TAM network relearn from b_{ji}, $w_{ji}(z_{ji})$ and $p_{jk}(o_{jk})$ of the fuzzy case of $I^*(W,Z)$ as the initial parameter values. After learning, the resultant fuzzy rule of the TAM network is exchanged to the existing fuzzy case in the fuzzy knowledge part.

4. EXAMPLE

In this section, we present an example of how to restructure fuzzy cases. Three data sets which include "△" and "•" are shown in Figure 6. 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 parameters of the fuzzy TAM network are set as follows:

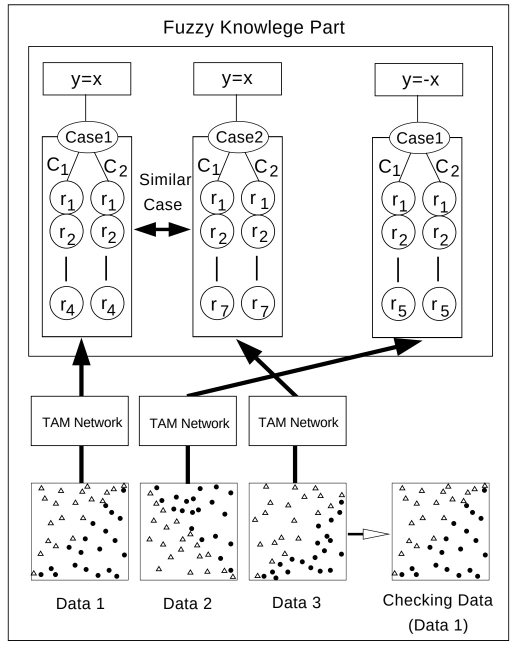

Figure 6: Procedure of the Fuzzy TAM Network

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

First, the fuzzy rule for the first data set is acquired from the TAM network and the fuzzy rule is memorized as a fuzzy case, "Case 1", of "Class($y = x$)" in the fuzzy knowledge part. Next, the fuzzy rule for the second data set is acquired from the TAM network and moved into the fuzzy knowledge part. The similarity measure of the second fuzzy rule with "Case 1" is obtained as $I^*(2,1) = 13.5$ as shown in Table 1. Since τ is set to 10.0, 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 fuzzy knowledge part, the similarity measures of the third fuzzy rule with the first fuzzy case and the second fuzzy case are calculated, respectively. Since $I(3,1) = -2.84$, $I(3,2) = -1.88$ and $I^*(3,1) = -1.88$, we let the TAM network relearn from b_{ji}, $w_{ji}(z_{ji})$ and $p_{jk}(o_{jk})$ of the first fuzzy case as the initial parameter values. The resultant fuzzy rule of TAM network is exchanged with the first fuzzy case and registered as "Case 1". The new fuzzy case, "Case 1", is shown in Figure 7.

To show the usefulness of the TAM network by restruc-

Table 1: Result of I(W,Z)

	Fuzzy Case 1 (in FKP)	Fuzzy Case 2 (in FKP)
Data 2	13.5	–
Data 3	-2.84	-1.88

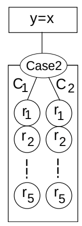

Figure 7: A New Fuzzy Case 1

turing fuzzy 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 2. 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 fuzzy TAM network is obtained as 60.0% and the number of categories is 5. The correct rate of the fuzzy TAM network is better than the TAM network and the number of categories is less than TAM network.

Table 2: Correct Rate and Categories

	Correct Rate		Categories	
	TAM	Fuzzy TAM	TAM	Fuzzy 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 fuzzy TAM network and showed the usefulness of fuzzy TAM network through an example. In the near future, we have to discuss the similarity measure how to compare fuzzy rules and formulate the algorithm.

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.

6. REFERENCES

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

[2] 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).

[3] 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).

[4] J.R.Williamson: Self-organization of Topographic Mixture Networks Using Attentional Feedback, *Neural Computation*, Vol.13, pp.563-593 (2001).

[5] 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).

[6] 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).

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

[8] B.Kosko : *Neural Networks and Fuzzy Systems, A Dynamical Approach to Machine Intelligence*, Prentice-Hall (1992).

[9] K.Nakaoka, T.Furuhashi and Y.Uchikawa: A Study on Apportionment of Credits of Fuzzy Classifier System for Knowledge Acquisition of Large Scale Systems, *Proc. of third IEEE Int. Conference on Fuzzy Systems*, Vol.2, pp.1797-1800 (1994).

[10] I.Hayashi, M.Umano, T.Maeda, 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, *Second International Conference on Knowledge-Based Electronic Systems* (*KES'98*), pp.69-78 (1998)

[11] I.Hayashi and J.R.Williamson: An Analysis of Aperture Problem Using Fuzzy Rules Acquired from TAM Networks, *IEEE International Conference on Fuzzy Systems* (*FUZZ-IEEE2002*), pp.914-919 (2002).