



Extraction of Knowledge from the Topographic Attentive Mapping Network and its Application in Skill Analysis of Table Tennis

by

Isao Hayashi¹, Masanori Fujii², Toshiyuki Maeda², Jasmin Leveille³, Tokio Tasaka⁴

The Topographic Attentive Mapping (TAM) network is a biologically-inspired classifier that bears similarities to the human visual system. In case of wrong classification during training, an attentional top-down signal modulates synaptic weights in intermediate layers to reduce the difference between the desired output and the classifier's output. When used in a TAM network, the proposed pruning algorithm improves classification accuracy and allows extracting knowledge as represented by the network structure. In this paper, sport technique evaluation of motion analysis modelled by the TAM network was discussed. The trajectory pattern of forehand strokes of table tennis players was analyzed with nine sensor markers attached to the right upper arm of players. With the TAM network, input attributes and technique rules were extracted in order to classify the skill level of players of table tennis from the sensor data. In addition, differences between the elite player, middle level player and beginner were clarified; furthermore, we discussed how to improve skills specific to table tennis from the view of data analysis.

Key words: Neural Networks, Fuzzy Logic, knowledge extraction, skill level, table tennis.

Introduction

In human motor skill research, movement skill is sometimes modelled as a hierarchical cerebellum model with feedback and feedforward functions that can adapt itself to environmental changes (Shiose et al., 2004). Kawato (1992) proposed the control model of Allen-Tsukahara as an internal model. When a difference exists between the desired trajectory and the realized trajectory of the movement, a difference signal is transmitted to Purkinje cells of the cerebellum and controls both movement output and starting timing. The Purkinje cells in the cerebellum control forward and inverse models for voluntary

movement. Based on this view of the cerebellum's contribution, in the present study, an internal model of the cerebellum is proposed in the form of a neural network acting through two kinds of processes, namely bottom-up processing of a signal flow to the integral representation of movement from a monofunctional layer, and top-down processing of the adjustment to the monofunctional layer from external observations.

The Topographic Attentive Mapping (TAM) network can be traced back to the ARTMAP family of models. The network structure consists of four layers: the feature layer,

¹ - Graduate School of Informatics, Kansai University, Takatsuki, Japan.

² - Faculty of Management Information, Hannan University, Matsubara, Japan.

³ - Wheatstone Technologies LLC, Washington D.C., U.S.A..

⁴ - Faculty of Health and Sports Science, Doshisha University, Kyotanabe, Japan.

basis layer, category layer and class layer. The basis layer connects the feature layer to either the category layer or the class layer via bidirectional projections. A node in the basis layer network combines the bottom-input signal propagated from the feature layer via excitatory synapses with the top-down feedback signal controlled from the output layer via inhibitory synapses. Upon presentation of a training pattern, if the network produces inaccurate output, the attentional top-down signal modulates the synaptic weights in the class and basis layers via winner-take-all learning in order to minimize the difference between the network output and the desired output. At the same time, nodes can be added incrementally to the category layer until the classification accuracy on the training set reaches a satisfactory level. However, such an addition of nodes can lead to overlearning of the training data and hence, impair generalization performance. In order to address this problem, we propose a network pruning algorithm that can remove unnecessary nodes and links based on fuzzy information entropy (Hayashi and Williamson, 2001), and we apply an ensemble learning model to the TAM network to be able to make a recognition rate high by assigning selection weight to misclassified data. In the pruning algorithm of the TAM network, each input variable one-by-one is extracted and the fuzzy information entropy of each attribute is calculated to determine its importance. The importance of each attribute is computed indirectly as a reduction in the recognition rate when the corresponding link and node are removed. Besides improving recognition accuracy, the main advantage of using the pruning method within a TAM network is that it leads to a natural way of extracting knowledge from the network structure. The integration of the pruning mechanism in a TAM network also bears strong similarities with earlier work on learning in the context of fuzzy logic processes (Hayashi and Umano, 1993; Hayashi et al., 1996; Kosko, 1992) as our pruning approach incorporates the metaphysical concept of fuzzy ID3 (Hayashi, 1996; Hayashi et al., 1999; Umano et al., 1994) into the TAM network structure, implying that the data is mined in the format of a fuzzy rule. Ensemble learning with AdaBoost (Dietterich, 2000; Freund, 1995; Schapire, 2003) is adopted here to combine

multiple TAM networks, each a weak classifier, and increase the recognition rate by differentially reweighting misclassified data. The final output is calculated with a majority rule after evaluating the multiple weak classifiers. According to this scheme, our algorithm may be considered as a form of a sensitive estimation method. Besides improving recognition accuracy, the main advantage of using the ensemble method within a TAM network is that it leads to a natural way of extracting a majority pattern from the data sets.

Input to our model takes the form of time-series extracted from motion picture data extracted from sensors (Kasai et al., 1994; Mochizuki et al., 2002). Perl and Baca (2003) employed the Kohonen Feature Map as a neural network for analysis of table tennis movement in order to characterize the strategic structure of table tennis. Parisi et al. (2015) proposed a learning system for providing feedback on a set of learned movements of powerlifting exercises captured with a Kinect device. The learning model was built upon a recursive extension of the Self-Organizing Map (SOM), MergeSOM. However, the proposed model did not aim to visualize a characteristic of the movement and produce a verbal description based on it. Therefore, it cannot be used to provide visualizations that would be useful to players. In our approach, we extract input attributes and skill rules of the forehand stroke of table tennis with the TAM network as an internal model (Hayashi et al., 2009; Maeda et al., 2014). The purpose of analyzing movement data with a TAM network is to visualize, for example, knowledge in the form of rules extracted from movement data. The aim of this research was to analyze the movement of players from observed data and help players to swing the racket based on acquired fuzzy rules. In order to do so, it is necessary to provide a verbal description of the observed data. A useful model of sports movements needs to provide for at least two kinds of capabilities: a high recognition rate and verbally-specified, easily understandable rules. Fuzzy rules based on languages can extract conjunctive variables with high membership values as important factors. This is a critical difference between fuzzy rules and methods based on statistical estimation. In this research, the technique rules are qualitatively specified as fuzzy rules from weights of network structure

and necessary attributes are estimated to distinguish the skill level.

Material and Methods

Participants

The presented experiment was performed in the Hannan University, Matsubara, Osaka, in 2007. Fifteen subjects were divided into three groups: elite players consisted of seven subjects who belonged to the table tennis club of the Hannan University, middle-level players were represented by three subjects who had previously belonged to a table tennis club during their junior high school or high school, and beginners consisted of five subjects without any table tennis experience.

Measures

Nine measurement markers were set to detect movement on the right upper arm of subjects, at 1) the acromioclavicular joint, 2) the acromion, 3) the head of radius, 4) the head of ulna, 5) the styloid process of radius, 6) the styloid process of ulna, 7) the right apex marker in the racket edge, 8) the left apex marker in the racket edge, and 9) the upper apex marker in the racket.

Procedures

A pitching machine (Yamato table tennis Co., Ltd., TSP2050) was set at about 30 cm from the end line of the table and diagonally from the subject. Balls were set to be thrown at elevations of 20 degrees, 25 speed levels and 30 pace levels. Subjects returned the ball in the area spanned by 75 cm inside from the end of the table to the opposite side. In order to trace the trajectory of a subject's movement, the forehand strokes were recorded for 10 min with a high-speed camera (Digimo Company, VCC-H300, resolution: 512×512 pixel, frame rate: 90 fps) placed 360 cm in front of the subject and at the height of 130 cm.

A 90 fps camera was used to record subjects for 10 min what resulted in 54000 recorded frames. From this data, image sequences of 40 to 120 frames were extracted. During each 10 min recording, the subject continued swinging the racket for 100 to 150 times. Overall, the trajectories generated by the elite subjects were very similar across swings. However, data from the elite trajectories could not be easily compared to that of the beginner and middle-level subjects due to variations in the start times in the latter's trajectories. In addition, the last swing trajectories

in the 10 min periods were often unreliable because of a lack of concentration at that time. Therefore, we used trajectory data for the middle portion of the 10 min recordings, e.g., in a 9 min 20 s recording, we took the swing at 280 s (9 min 20 s = 560 s, divided by two equals 280 s). In addition, the start time of the racket swing was extracted from the time when the position of the take-back was minimized in the x direction to the time when the position of the follow-through became greatest in the x direction. Therefore, from one experimental sequence consisting of swing movements for ten minutes, it was possible to extract between 40 and 120 frames of data. As a result, the number of training data points for seven subjects was approximately 600 and the number of testing data points for two subjects was approximately 200. In each image frame, two-dimensional (x, y) coordinates of nine measurement markers relative to the original position of the subject's shoulder on the first frame were obtained. The observed positions of markers are presented in Figure 1 and the speed of the horizontal direction (x) in Figure 2.

The following observations can be made from Figures 1 and 2:

- When comparing two elite players, the coordinates of positions from M_1 to M_9 were close to each other. The correlation coefficients were evaluated at 0.985 in the x direction and 0.79 in the y direction, respectively. Elite players thus acquired a fairly stable racket swinging motion.
- The swing pattern of elite players allowed them to reach maximum speed at the point of contact with the ball.
- The M_1 to M_9 coordinate positions for two of the middle-level players were somewhat less stable: the correlation coefficients obtained were 0.919 in the x direction and 0.607 in the y direction. Overall, the trajectories did not form a smooth, oval-shaped forehand drive.
- The swing speeds of middle-level players for markers M_7 and M_9 have two peaks and appear to be adjusted at the moment of impact with the ball.
- When looking at the data of the three beginners, the coordinates of positions M_1 to M_9 varied considerably across players. The correlation coefficients were evaluated

at 0.0703 in the x direction and -0.04 in the y direction. Thus, there is no single way to characterize the beginners' swing pattern. Also, it may be noted that the shoulder (M_1) moved more than in elite and middle-level players.

- Looking at the swing speed of beginners at marker M_9 , the speed appears to be reduced just before hitting the ball, and subjects seem to wait until the ball hits the racket. This is sometimes referred to as either "a movement to meet a ball with the racket" or "a movement to delay the body". Non-zero speeds are recorded at markers M_1 and M_4 even though the speeds at M_7 and M_9 in the same image frames are both zero.
- Regarding the width of the swing pattern in the horizontal direction (x) at the first (M_1), fourth (M_4) and ninth (M_9) markers, elite players appear to swing rather compactly, whereas beginners make large swings.

Statistical Analysis

Internal Model

In the analysis of sport skills, body motion is typically measured with electromyography, in which action potentials are recorded when muscular fibers are excited during movements. However, we used measurement markers attached to the body to record various limb positions instead. In this paper, we introduced a neural network as an internal model in which a mono-functional layer generated a single function, while a meta-layer adapted to environmental changes. Using the TAM network as a kind of the internal model, table tennis skills were characterized using both the trajectory data of forehand strokes and the coach's evaluation of corresponding table tennis skills.

TAM Network

It was assumed that observation data of the R unit were given in M inputs and one output as data set D . The s -th data of the i -th input variable was denoted as v_{si} , $s = 1, 2, \dots, R$, and the output data was represented by O_s .

Each data point was rank-ordered in the feature layer. In the i -th input feature, the input data of the R unit was sorted in ascending order again and the input data, I_{si} , was normalized. Using the input data I_{si} , the distributed data in the feature layer was provided as f_{sih} .

$$I_{si} = \frac{s - 0.5}{R}, \quad i = 1, 2, \dots, M. \quad (1)$$

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

where L is the number of distributed data samples, and h is a sample-specific suffix, $h = 1, 2, \dots, L$. f_{sih} can be reduced to f_{ih} since only one input sample was included at a time.

Figure 3 shows the structure of the TAM network. The activity value, x_{ji} , of each node in the basis layer was calculated by taking into account the excitatory synapse weight, w_{jih} , of the projections from the feature layer and the inhibitory synapse weight, b_{ji} , from the class layer multiplied by the vigilance parameter ρ . The output, y_j , from category node j was then calculated as follows:

$$y_j = \prod_{i=1}^M x_{ji} = \prod_{i=1}^M \frac{\sum_{h=1}^L f_{ih} w_{jih}}{1 + \rho^2 b_{ji}}. \quad (3)$$

The output of the TAM network was in turn given by the index of the maximally active unit in the class layer:

$$K = \{k \mid \max_k z_k\} = \{k \mid \max_k \sum_{j=1}^N y_j p_{jk}\} \quad (4)$$

where p_{jk} , $k = 1, 2, \dots, U$ is a synaptic weight between category node j and class node k .

Let K^* denote the desired output of the class layer for a given input vector in a training dataset. If the output K of the TAM network does not correspond to the desired output class K^* , an "attentional" mechanism is invoked whereby the vigilance parameter ρ increases until either $z_{K^*}/z_K \geq OC$, where OC is a threshold or the maximal vigilance level $\rho^{(max)}$ is reached, that is:

$$\text{While } z_{K^*} / z_K < OC \text{ and } \rho < \rho^{(max)}$$

$$(a) \rho = \rho + \rho^{(step)}$$

$$(b) \text{ equation (3) and (4)}$$

If the vigilance parameter ρ reaches its maximum level, $\rho^{(max)}$, one new node is added to the category layer. However, if the constraint $z_{K^*}/z_K \geq OC$ is satisfied, weight adaptation occurs using a feedback signal, y_j^* , from the class layer to the category layer, computed as follows:

$$z_k^* = \begin{cases} 1 & ; \text{ if } k = K^* \\ 0 & ; \text{ otherwise} \end{cases} \quad (5)$$

$$y_j^* = \frac{\prod_{i=1}^M x_{ji} \times \sum_{k=1}^U z_k^* p_{jk}}{\sum_{j=1}^N \prod_{i=1}^M x_{ji} \times \sum_{k=1}^U z_k^* p_{j'k}} \quad (6)$$

The feedback signal is then used to govern learning:

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

$$\Delta p_{jk} = p_j^{(rate)} y_j^* (z_k^* - p_{jk}) \quad (8)$$

$$\Delta w_{jih} = w_j^{(rate)} y_j^* (f_{ih} - w_{jih}) \quad (9)$$

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

$$w_j^{(rate)} = \frac{\alpha}{\alpha \beta(M) + n_j} \quad (11)$$

where

$$\beta(M) = \frac{\lambda^{1/M}}{1 - \lambda^{1/M}}, \quad \lambda \in (0, 1) \quad (12)$$

$$\Delta n_j = \alpha y_j^* (1 - n_j) \quad (13)$$

and α , λ and $b_j^{(rate)}$ are constant parameters. Parameter $p_j^{(rate)}$ acts in a way similar to the revision parameter in simulated annealing and $w_j^{(rate)}$ is the revised value of the bias $\beta(M)$ of the M -dimensional inputs.

In the training phase, learning of w_{jih} , p_{jk} and b_{ji} proceeds upon presentation of each input datum. Each presentation of the whole training set is called an epoch and training consists of multiple epochs. After learning is completed, the values of w_{jih} , p_{jk} and b_{ji} should be close to f_{ih} , z_k^* and x_{ji} , respectively, due to winner-takes-all and adaptive learning (Carpenter et al., 1991).

The usefulness of the TAM network in various contexts was demonstrated by Williamson (2001). However, one issue of the learning algorithm in TAM networks is that the unconstrained addition of nodes in the network can lead to severe overlearning, thereby limiting the generalization capability of the model. Furthermore, it would be advantageous to be able to reduce the number of features learned, i.e., to minimize the number of input variables while maintaining accuracy. In order to address both problems, we proposed a new algorithm for pruning unnecessary nodes and links after the node addition step in the basic learning algorithm.

In order to introduce the pruning algorithm, it is useful to first summarize the structure of the TAM network. The four layers of the TAM network can be divided functionally into two groups: a lower level consisting of the feature and basis layers, and an upper level comprising the category and class layers. In the lower level, the synaptic weight w_{ji} of the i -th basis node in the j -th

category node should be similar to the data f_i of the i -th feature due to winner-takes-all learning. Therefore, the category node effectively encodes the distribution of the input data in the synaptic weights of the basis layer. On the other hand, the synaptic weight p_{jk} between the j -th category node and the k -th class node in the upper level represents the proportion in which the j -th category node contributes to the k -th class node. By taking w_{ji} as a membership function, equation 14 shows that it is possible to interpret category node r_j as the j -th rule of a set of fuzzy rules:

$$\left. \begin{array}{l} r_1 : \text{if } f_1 \text{ is } w_{11} \text{ and } \dots \text{ and } f_M \text{ is } w_{1M} \\ \quad \text{then } C_1 = p_{11}, \dots, C_U = p_{1U} \\ \vdots \\ r_j : \text{if } f_1 \text{ is } w_{j1} \text{ and } \dots \text{ and } f_M \text{ is } w_{jM} \\ \quad \text{then } C_1 = p_{j1}, \dots, C_U = p_{jU} \\ \vdots \\ r_N : \text{if } f_1 \text{ is } w_{N1} \text{ and } \dots \text{ and } f_M \text{ is } w_{NM} \\ \quad \text{then } C_1 = p_{N1}, \dots, C_U = p_{NU} \end{array} \right\} \quad (14)$$

where C_k , $k = 1, 2, \dots, U$ stands for the output of the k -th class node, and r_j , $j = 1, 2, \dots, N$, that of the j -th category node.

From equation 14, it is possible to extract fuzzy rules from the structure of the TAM network. It can be said that an extracted fuzzy rule can express the relationship between input and output data pairs as represented by a network structure of minimal size following pruning.

Umamo et al. (1994) proposed fuzzy ID3 to obtain fuzzy rules from given input and output data pairs. In fuzzy ID3, one variable among the input variables is selected and the fuzzy information entropy is calculated. The input variable with the maximal fuzzy information entropy is deemed the most important and is thereby included into a fuzzy decision tree. A fuzzy rule that captures the relationship between input and output data pairs is obtained by repeating this selection process. The selection process was used as part of the pruning algorithm.

In the pruning algorithm that we propose, the input variable that maximizes the fuzzy information entropy is extracted consecutively from a set of input variables in the same manner as in fuzzy ID3. Let I^* be the set of features extracted from the entire input feature set when

fuzzy information entropy is maximized. Making use of I^* , for the s -th data point in data set D , the output of the j -th category node is expressed as:

$$\delta_{js} = \prod_{i \in I^*} x_{jis} . \quad (15)$$

Activity ρ_{js} can be interpreted as a membership function of the antecedent part of the j -th fuzzy rule for the s -th data point since equation 14 is considered as a fuzzy rule.

In the fuzzy ID3 algorithm, whenever the i -th attribute is selected, the conditional probability $P(k | i)$ of the k -th class given the i -th attribute is first computed and used to determine whether that attribute should be an attribute of the decision tree. Instead of selecting the i -th attribute and calculating the conditional probability $P(k | i)$, in our pruning algorithm, feature i was added to feature set I^* and the probability G_{jk} of a data having k as an output class was calculated:

$$G_{jk} = \frac{\sum_{s \in \Psi_k} \gamma_{js} \times p_{jk}}{\sum_{s=1}^R \gamma_{js} \times p_{jk}} \quad (16)$$

where by equation 15,

$$\gamma_{js} = x_{jis} \times \delta_{js} \quad (17)$$

and Ψ_k is the subset of inputs with the output class k .

We defined fuzzy information entropy when feature i was added to I^* as follows:

$$H(i) = - \sum_{j=1}^N g_j \sum_{k=1}^U G_{jk} \log_2 G_{jk} \quad (18)$$

where prior probability g_j in category node j was calculated as follows.

$$g_j = \frac{\sum_{s=1}^R x_{jis}}{\sum_{j=1}^N \sum_{s=1}^R x_{jis}} \quad (19)$$

The feature that maximized the fuzzy information entropy was added to feature set I^* and was denoted as i^* :

$$i^* = \{i \mid \max_i H(i)\} . \quad (20)$$

Since necessary features are selected sequentially in equation 20, the order of selection of a feature offers a means to define the importance of the attribute. Feature deletion also proceeds sequentially by applying the three following strategies to prune unnecessary links and nodes:

- For each category node, the strength of the combination of each class of the class layer

with that category node is estimated and unnecessary combinations are removed.

- For each category node, the strength of the combination of each feature of the feature layer with that category node is estimated and unnecessary combinations are removed.
- For each class node, the strength of the combination of each category node in the category layer with that class node is estimated and unnecessary combinations are removed.

The following three pruning rules were formulated based on the above strategies.

[Pruning Rule 1]

Link $j - k'$, $k' = 1, 2, \dots, U$, $k' \neq k$, between category node j and class node k' is removed if the condition in equation 21 is satisfied for testing data in the j -th category node.

$$G_{jk} \geq \eta \quad (21)$$

where η is a threshold. The link $j - i'$ with the remaining i' -th feature node that is not included in the feature set I^* , that is, $i' \notin I^*$ and the j -th category node are removed at the same time.

[Feature 1]

If the feature i'' selected by equation 20 satisfies condition 22, then equation 21 is necessarily satisfied for a specific j, k .

$$\sum_{s \in \phi_k} x_{ji''s} \geq \sum_{s \in \phi_k} x_{ji''s} \quad (22)$$

[Pruning Rule 2]

The link $j - i$ between the j -th category node and the i -th feature node is removed if condition 23 is satisfied for testing data in the j -th category node.

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

where θ is a threshold. The link $j - i'$ between the j -th category node and the i' -th feature node, that is $i' \notin I^*$, is removed at the same time.

[Feature 2]

If the j -th category node satisfies equation 23 once, then condition 23 is necessarily satisfied in the next pruning iteration I^* , due to the following assumption.

$$x_{jis} \leq 1 \text{ and } \prod_{i \in I} x_{jis} \geq \prod_{i \in I^*} x_{jis} . \quad (24)$$

Then, in connection with class k and category j , we defined the ratio of the activity of

the j -th category node to the activity all category nodes as φ_{jK} :

$$\varphi_{jK} = \frac{\sum_{s \in \Gamma_K} \gamma_{js'} \times p_{jK}}{\sum_{j=1}^N \sum_{s \in \Gamma_K} \gamma_{js} \times p_{jK}} \quad (25)$$

where $\Gamma_K = \{s \mid K=K^*, K = \max_k \sum_{j=1}^N y_j p_{jk}\}$.

Note that φ_{jK} expresses the importance of each category node connected to class K .

[Pruning Rule 3]

Link $K - j'$, along with the K -th class node, the most active node in the class layer, and the j' -th category node except category node j , that is, $j', j' = 1, 2, \dots, N, j' \neq j$, is removed if the condition 26 is satisfied for testing data in the K -th class node:

$$\varphi_{jK} \geq \xi \quad (26)$$

where ξ is a threshold.

[Feature 3]

The output K of the TAM network is equal to the correct output K^* as long as the following condition concerning threshold ξ is satisfied:

$$\xi \geq \max_{k \neq K} \frac{\sum_{s \in \Gamma_k} \sum_{j=1}^N p_{jk}}{\sum_{s \in \Gamma_k} \sum_{j=1}^N y_{js} p_{jk}} \quad (27)$$

The pruning algorithm is summarized below:

- [Step 1] Extract arbitrary feature i and calculate the fuzzy information entropy $H(i)$ of equation 18 over the testing data.
- [Step 2] Select a feature that maximizes the fuzzy information entropy and add it to feature set I^* . Denote $I^* = \{i^*\}$, where $i^* = \{i \mid \max_i H(i)\}$.
- [Step 3] As per the first pruning rule, remove link $j - k', k' = 1, 2, \dots, U, k' \neq k$, between the j -th category node and the k' -th class node if the condition 21 is satisfied for testing data in the j -th category node. Links $j - i'$ with the remaining i' -th feature node that are not included in the feature set I^* , $i' \notin I^*$ and the j -th category node are removed at the same time.
- [Step 4] Following the second pruning rule, link $j - i$ between the j -th category node and the i -th feature node is removed if condition 23 is satisfied for testing data in the j -th category node. The link $j - i'$ between the j -th category node and the i' -th feature node, $i' \notin I^*$, is removed at the same time.

[Step 5] Based on the third pruning rule, link $K - j'$ along with the K -th class node, the most active node in the class layer, and the j' -th category node, that is, $j', j' = 1, 2, \dots, N, j' \neq j$, are removed if the condition 26 is satisfied for testing data in the K -th class node.

[Step 6] Unnecessary links and nodes are removed. In addition, features with unnecessary links are removed.

[Step 7] Repeat steps 1-6 until all features are included in feature set I^* .

The pruning algorithm yields a neural network with a minimal number of links and nodes. Since we can calculate the importance of each feature and eliminate unnecessary features, the pruning algorithm can be considered a feature reduction method.

Adaboost Type TAM Network

Adaboost is an outstanding boosting method. In each iteration of the Adaboost algorithm, TRD is selected from the set of misclassified data with higher weights than 50% and then these data are applied to a weak classifier in the consecutive iteration. After the weak classifier is identified, the weights of the data are updated. The procedure is repeated until a maximum number of iterations is reached or until the current recognition rate of CHD is higher than the previous recognition rate. The joint output is calculated by majority rule decision of the multiple weak classifiers $M_1, M_2, \dots, M_i, M_L$ when CHD is given to these models.

Results

In the present study, the data of the two-dimensional (x, y) coordinates of all nine markers were analyzed with the TAM network. Since table tennis skills are better characterized with time-series, we created data sets by concatenating data across five consecutive frames. Another reason to use multiple frames as input was to be able to arbitrarily change the order presentation of each input variable to the TAM network. If the data had been input in the form of single frames, it would have not been possible to describe a series of racket swings. Therefore, past frames were also provided to the network in order to describe the series of swing movements. This is a method commonly used in Autoregressive Models (Pandit and Wu, 1983). In addition, our best recognition results were obtained when using series of 5

frames. The output of the model was the skill evaluation discretized to three levels. As a result, a data set consisted of 90 input variables (x and y coordinates of nine markers over five frames, that is $2 \times 9 \times 5$) and three classes.

The training data (TRD) consisted of three kinds of players, i.e., two elite players who were selected from three elite players, two middle level players and two beginners who were selected from four beginners, and the testing data (CHD), which was constituted from the data of one elite player and one beginner. The results were strongly dependent on which kind of data was used during training and testing. Therefore, for beginner subjects, the correlation coefficient of the position coordinate was calculated at each marker and a data set D which included subjects with the highest correlation coefficient among four beginners in TRD and CHD was created, respectively. If data had been extracted randomly from all swing movements across all players, we would not have been able to define the racket motion because part of the movement might be missing. The number of frames in a swing from the start position to the end position for each player was different as the time for each swing movement was different for beginners, middle and elite players. Therefore, two players for the beginner and middle level and three for the elite level were used so that the amount of data for each level was approximately the same. Using the two extremes (elite and beginner) does not fully test the TAM. It is likely that these two categories could easily be separated by using simple rules. A more balanced and systematic approach would be to put aside one elite, one middle level and one beginner player for testing and the rest of athletes for training. All other participants would be then assigned to combinations to the testing group and the rest to training. This way all participants would become part of the testing group in all combinations with others, and they would be also part of the training group with others. Such a bootstrap method would provide an unbiased analysis.

The recognition rate obtained with the TAM network for TRD was 53.7%, and the recognition rate for CHD was 57.5%. This low accuracy could in part be due to a difference in the number of observations for each class. Therefore, for data set D, we constituted datasets by adding four

consecutive frames to each initial input frame to get 5-frame long sequences and let the number of datasets increase by the added amount of data. The result is shown in Table 1. TAM (D) means the recognition rate of data set D, and TAM (D+) shows the recognition rate of the revised data set D with added data. Simultaneously, the recognition rates of C4.5, Native Bayes Tree (NBT), Random Forest (RF) are shown for comparing the TAM network with their data mining methods for data set D. In particular, C4.5 is an algorithm used to generate a decision tree developed by Quinlan (1993) and is an extension of the ID3 algorithm.

The recognition rate for the TAM network using dataset D+ improved compared to when using D. The recognition rates of NBT and RF as to TRD equalled 100%, and that of C4.5 was 98.1%, which could be explained as overfitting given the corresponding low accuracy of testing data. The differences in recognition rates across training and testing sets were 67.2% and 74.6% for NBT and RF, respectively. Correspondingly, the difference between training and testing sets for C4.5 was 54.8%. This suggests again that NBT, RF and C4.5 overfitted the training data. On the other hand, differences in recognition rates between training and testing sets for the TAM(D+) and TAM(D) network were rather small, at the level of 18.4% and -3.8%, respectively. This suggests the TAM network did not overfit the training data. However, because the TAM network did not achieve high recognition rates on either the training or testing set, we re-trained the TAM network with Adaboost and showed the results in the next session.

Next, the sensitivity of markers was analyzed using a feature selection method. The selection of variables should be based on increases and decreases in the recognition rate. In general, the recognition rate decreases when reducing the number of variables. However, when unnecessary variables are removed this decrease is not significant. Conversely, when important variables are removed the recognition rate decreases greatly. Therefore, we should note the rate of increases and decreases in the recognition rate rather than its value. In the results of the experiment on variable selection, D+ was used as the data set. When all nine markers were used, the number of input variables was 90. That number

decreased to 70 inputs when the number of the markers was decreased from nine to seven. In particular, we should pay attention to the variable of which the recognition rate decreased when reducing variables. The resulting recognition rates when removing markers are shown in Table 2. We showed that the average recognition rate was 10 times greater compared to the one obtained for TRD. When M1 and M2 were temporarily removed, the recognition rate of the TAM network decreased from 61.2% to 42.9% reaching the lowest value. Therefore, we concluded that the most important markers were M1 and M2. By the same procedure, inputs were sorted in importance as follows: M1, M2 → M7, M8, M9 → M5, M6 → M3, M4. The recognition rate was conversely increasing when M5, M6 and M3, M4 were removed. In addition, since pairs of markers attached to the same bones were located in the front side and back side of the bone, it was appropriate to delete pairs of markers at the same time. In fact, the number of markers could be reduced from 9 to 2 or 5, as pairs of markers showed similar movements.

Lastly, fuzzy rules that characterized table tennis skills were extracted. The TAM network consists of four layers of hierarchical structure. The feature and basic levels represent a mono-functional mechanism, while the category and class levels represent a meta-concept. Using the structure of the TAM network, the relationship between the mono-functional skill and the meta-skill could be determined as a fuzzy rule.

Therefore, the J -th category node was first selected, where p_{jk} became the maximum at each class node in the data set D^+ under the condition of equation 21:

$$J = \{j \mid \max_j p_{jk}, \quad k = 1, 2, 3\}. \quad (28)$$

In addition, we calculated the real value w_{ji} of the J -th category node for each input as a singleton membership function:

$$w_{ji} = \frac{\sum_{h=1}^L w_{jih}}{L}, \quad \text{for } \forall i. \quad (29)$$

The set of linkages represented a fuzzy rule when we extracted linkages where w_{ji} represented the maximum for each player. To express fuzzy rules for elite and beginner players, five categories with high p_{jk} were extracted for

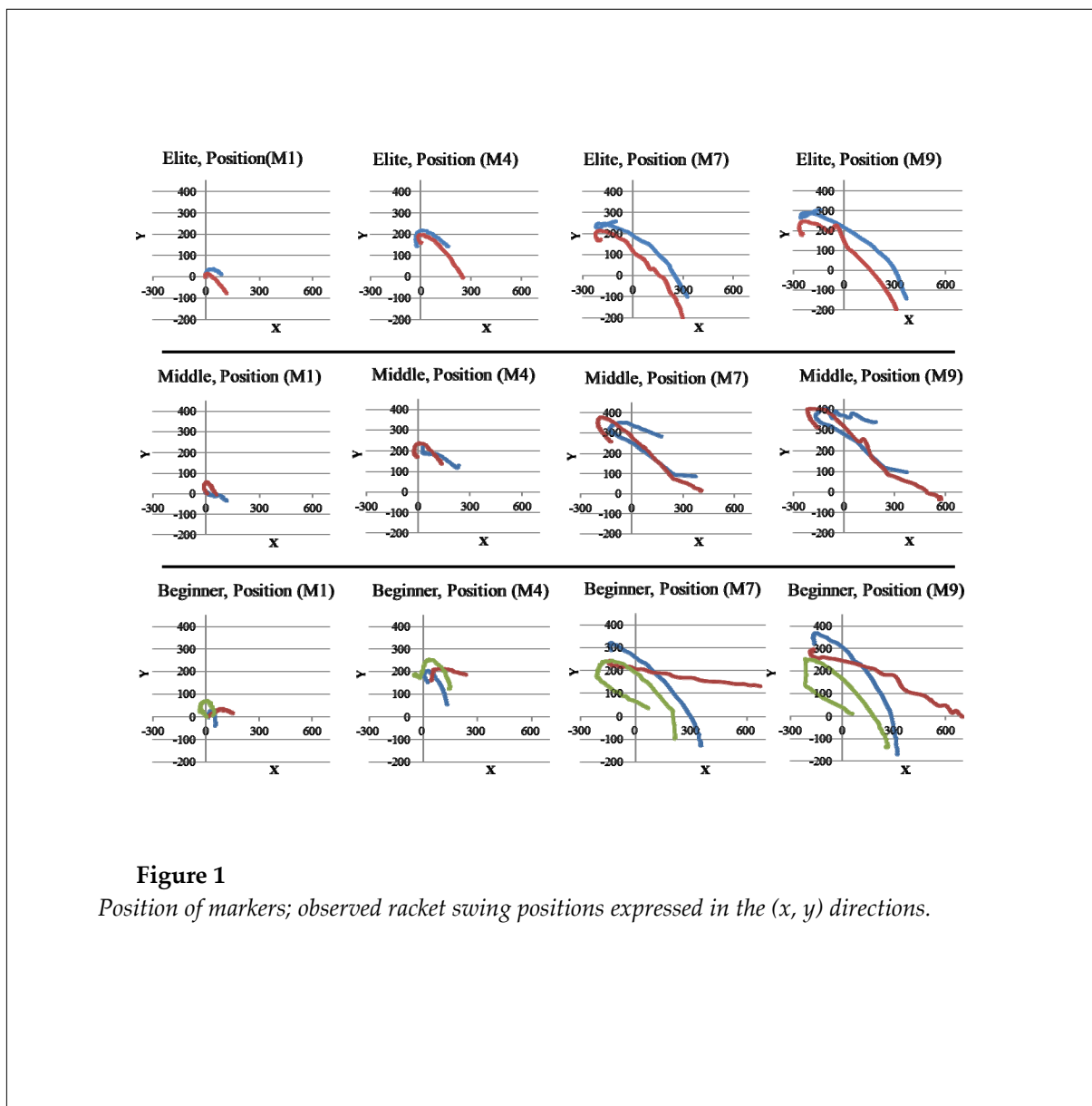
each class. Since a category expresses a rule, these categories expressed the rules of the highest five p_{jk} 's. In the first rule, w_{ji} of the marker on the racket in the horizontal direction was high, implying that this movement of the racket was important. The value of w_{ji} of each marker in the vertical direction was low, meaning that the swing movement was stable and without rough wavy movements. In both horizontal and vertical directions after the third rule, the value of w_{ji} was constant and the swing was stable. On the other hand, the vertical rise and fall motion was a salient feature of beginners' motion patterns and was extracted as the first rule. The horizontal movement at the shoulder and the elbow was remarkable in the second rule, implying "a movement to delay the body". Figure 4 shows the first rule of the elite player and beginner. In this Figure, the difference between the elite player and the beginner is shown very conspicuously, making it easy to formulate table tennis skills as a rule.

The recognition rate achieved with the TAM network was better than that obtained with other data mining methods. However, the recognition rate was still fairly low. Therefore, we applied the Adaboost algorithm, which is a kind of an ensemble learning model, to the TAM network in order to improve its recognition rate. First, a dataset was built so that the number of data samples of each dataset became approximately the same across classes. By adjusting the dataset in this way, the number of elite players in TRD of the data set D^{++} became 78, the number of middle level players became 73 and the number of beginners became 98. On the other hand, as to CHD of data set D^{++} , the number of elite players and beginners was 54 and 40, respectively. The recognition rate of TRD was 61.2%, while the recognition rate of CHD was 43.0%. The recognition rate was computed as the average across 10 repetitions of the experiment. For dataset D^{++} , three kinds of data groups were constructed as Adaboost was by default defined for two-class problems. That is, dataset D^{++} was partitioned into three groups with two classes, i.e., D1 which included the beginners and the others, D2 that comprised the middle level players and the others, and finally D3 composed of the elite players and the others. These data sets were analyzed by the Adaboost type TAM network

with epoch = 3, $\alpha = 0.0000001$, $\lambda = 0.33$.

The results are summarized in Table 3. The average recognition rate was 10 times of the data sets. As to TRD of dataset D1, the Adaboost algorithm was repeated three times, and 149 data were selected as misclassified TRD at the first step of the algorithm, and 42 data were selected as misclassified TRD at the second step. In the same way as to the data set D2, 149 data were selected as misclassified TRD in the first step of the algorithm, and 44 data were selected as misclassified TRD in the second step. As to the

data set D3, 149 data were selected as misclassified TRD in the first step, and 42 data were selected as misclassified TRD in the second step. The average recognition rate of the Adaboost-type TAM network for TRD improved to 73.1%, which was better than 69.3% of the original TAM network. As to CHD, the recognition rate of Adaboost type TAM network increased to 65.5% as the recognition rate of the original TAM network was 58.2%. A t-test applied to these results showed that Adaboost-type training of the TAM network improved accuracy over the original TAM network ($p = 0.014$).



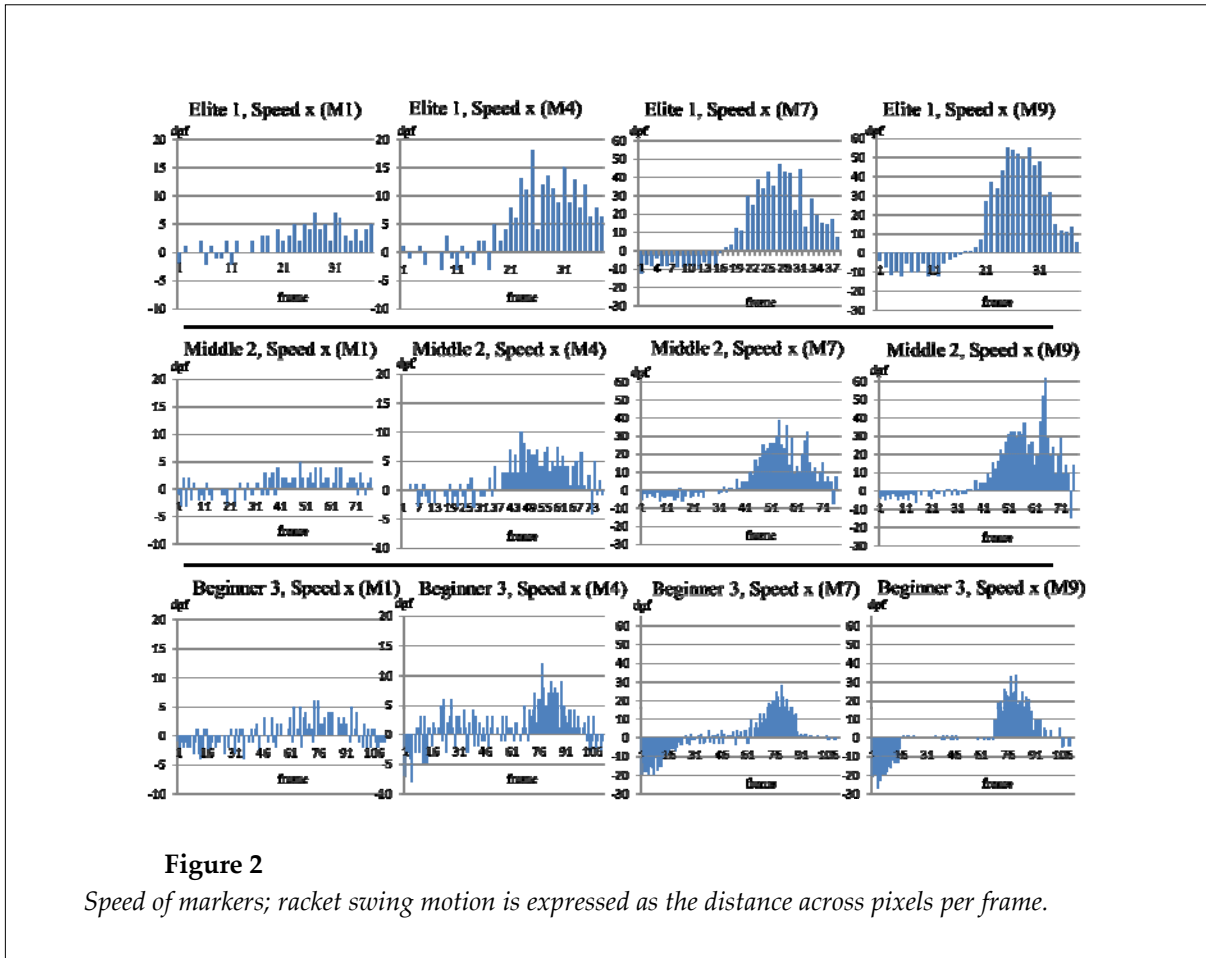


Figure 2

Speed of markers; racket swing motion is expressed as the distance across pixels per frame.

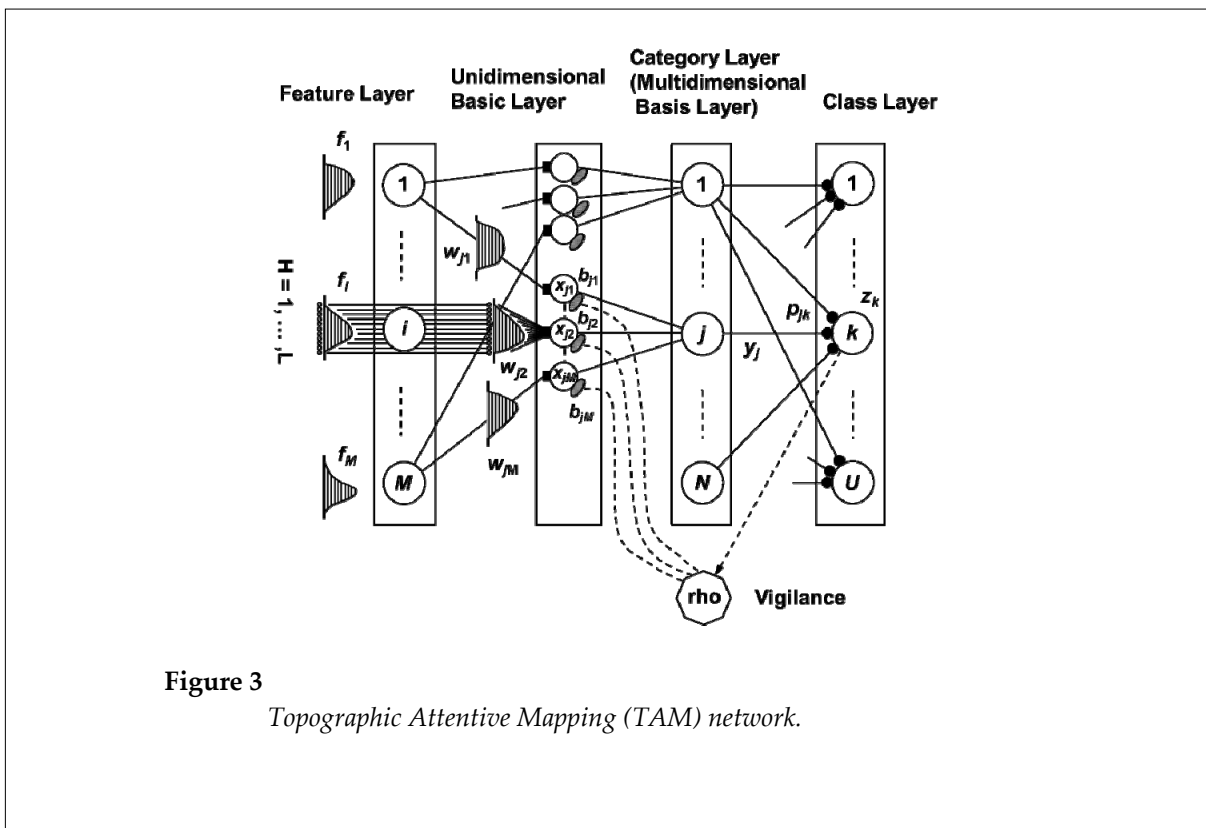


Figure 3

Topographic Attentive Mapping (TAM) network.

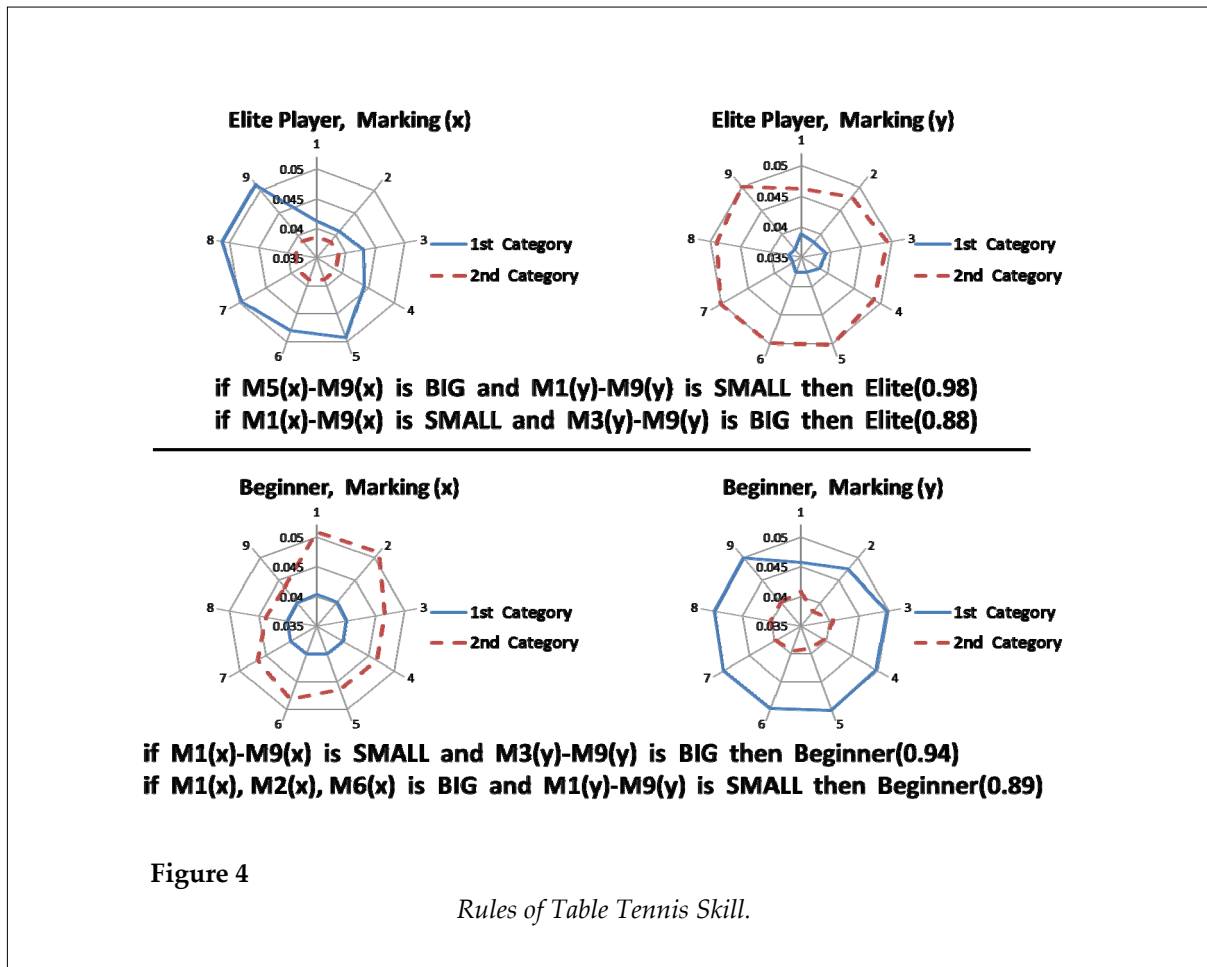


Table 1

Recognition Rates of Revised Data Sets:

Methods	Recognition Rate (%)			
	TRD	CHD	Average	Subtraction
TAM(D+)	61.2	43.0	52.1	18.2
TAM(D)	53.7	57.5	55.6	-3.8
C4.5	98.1	43.3	70.7	54.8
NBT	100.0	32.8	66.4	67.2
RF	100.0	25.4	62.7	74.6

Table 2

Sensitivity of Input Variables.

Number of Input Variables	Omitted Input Variables and Recognition Rate (%)								Minimized Recognition Rate (%)
	M1, M2		M3, M4		M5, M6		M7, M8, M9		
18	61.2		61.2		61.2		61.2		61.2
12-14	42.9	-18.3	57.4	-3.8	51.1	-10.1	48.2	-13.0	42.9
8-10			45.9	+3.0	48.4	+5.5	41.6	-1.3	41.6
4			42.9	+1.3	42.0	+0.4			42.0
Priority Order	1		4		3		2		

Table 3

Recognition Rate by the Adaboost Type TAM Network

Data	Data Sets	Recognition Rate of TAM Network (%)	Recognition Rate of Adaboost Type TAM Network (%)					Majority Result
			M ₁	M ₂	M ₃	Average		
TRD	D ₁	67.0	67.0	70.2	75.0	70.7	-	
	D ₂	71.0	71.0	74.0	80.6	75.2	-	
	D ₃	70.0	70.0	72.7	77.5	73.4	-	
	Average	69.3	69.3	72.3	77.7	73.1	-	
CHD	D ₁	58.5	58.5	64.6	(56.9)	61.6	58.5	
	D ₂	58.0	58.0	69.0	(42.0)	63.5	69.0	
	D ₃	58.0	58.0	69.0	(42.0)	63.5	69.0	
	Average	58.2	58.2	67.5	47.0	62.9	65.5	

Discussion

In this section, we will discuss the evaluations of the model. First, the following observations with regard to the coordinates of positions can be made from Figures 1 and 2:

- When comparing two elite players, the coordinates of positions from M1 to M9 were close to each other. Elite players thus acquired a fairly stable racket swinging motion.

- The M1 to M9 coordinate positions for two of the middle-level players were somewhat less stable. Overall, the trajectories did not form a smooth, oval-shaped forehand drive.
- When looking at the data of the three beginners, the coordinates of positions M1 to M9 varied considerably across players. There is no single way to characterize the beginners' swing pattern.

Since the correlation coefficients of the horizontal direction (x) of the markers for elite

and middle level players were higher than 0.9, it can be inferred that the swing movements in the horizontal direction of each elite and middle level player were very similar. At the same time, the correlation coefficient for the vertical direction (y) was higher than 0.6, and thus, the movement in the vertical direction was also similar. However, the correlation coefficients for beginners were less than 0.1 in both horizontal (x) and vertical (y) directions, and consequently, the movements of the beginners were different from each other. The recognition accuracy of NBT, RF and C4.5 on the training set was very high (Table 1), which can be explained as overfitting given the corresponding low accuracy on the test data. Since the beginner category is hard to clearly define, it is difficult to obtain high recognition accuracy using a discriminatively trained model. Differences in recognition rates across training and test sets were very large for NBT, RF and C4.5. This suggests again that NBT, RF and C4.5 overfitted the training data. On the other hand, differences in recognition rates between training and testing sets for the TAM(D+) and TAM(D) network were rather small. This suggests that the TAM network did not overfit the training data.

Next, the following observations with regard to the racket swing motion can be made from Figures 1 and 2:

- Elite players acquired a fairly stable racket swinging motion.
- The trajectories of the middle-level players did not form a smooth, oval-shaped forehand drive.
- The shoulder (M1) of the three beginners moved more than in elite and middle-level players.

Using a feature selection method, the sensitivity of markers was analyzed. We concluded that the most important markers were M1 and M2, and inputs were sorted in importance as follows: M1, M2 \rightarrow M7, M8, M9 \rightarrow M5, M6 \rightarrow M3, M4. Based on these results, it could be stated that important variables to evaluate player's ability are firstly, 1) the acromioclavicular joint and 2) the acromion, and secondly, markers 7 to 9 located on the racket. These results are consistent with the conclusions reached from Figures 1 and 2.

Lastly, the following observations with regard to the swing speed can be made from

Figures 1 and 2:

- The swing pattern of elite players allows them to reach maximum speed at the point of contact with the ball.
- The swing speeds of middle-level players for markers M7 and M9 have two peaks and appear to be adjusted at the moment of impact with the ball.
- Looking at the swing speed of beginners at marker M9, it appears to be reduced just before hitting the ball, and subjects seem to wait until the ball hits the racket. This is sometimes referred to as either "a movement to meet a ball with the racket" or "a movement to delay the body". Non-zero speeds are recorded at markers M1 and M4, even though the speeds at M7 and M9 in the same image frames are both zero.

Figure 4 shows the first rule obtained for both an elite player and a beginner. In the first rule of the elite players, the weight value of the racket in the horizontal direction was high, implying that this movement of the racket was important. The weight value of each marker in the vertical direction was low, meaning that the swing movement was stable and without rough wavy movements:

Elite player: If M5(x) to M9(x) are BIG and M1(y) to M9(y) are SMALL then the degree of elite player is 0.98.

On the other hand, the vertical rise and fall motion was a salient feature of beginners' motion patterns and thus, it was extracted as the first rule.

Beginner : If M5(x) to M9(x) are SMALL and M3(y) to M9(y) are BIG then the degree of beginner is 0.94.

From the TAM network, regarding the width of the swing pattern in the horizontal direction (x), elite players appear to swing rather compactly, whereas beginners make large swings. In addition, the difference between the elite player and the beginner is shown very conspicuously, making it easy to formulate table tennis skills as a rule. In particular, when a player hit a ball with the racket while forming a straight line with his head, shoulder and foot at the center of his body to create a stable axis of rotation, the shoulder of the beginner did not remain stable. The swings of the elbow, wrist and racket were also wide in the

vertical direction. The horizontal movement at the shoulder and the elbow was remarkable in the second rule, implying either “a movement to meet a ball with the racket” or “a movement to delay the body”.

Fuzzy rules based on languages can extract conjunctive variables with high membership values as important factors. As a result, with the TAM network, input attributes and technique rules were extracted in order to classify the skill level of players of table tennis from the sensor data. However, we should explore the structure of the resulting internal

model in order to better understand how to improve table tennis skills in the future.

Conclusion

In this paper, a dataset of forehand strokes of table tennis was analyzed with a TAM network and a boosted TAM network. In addition, fuzzy rules describing the skill of forehand strokes were extracted depending on the player’s performance level. Future work should include exploring the structure of the resulting internal model in order to better understand how to improve table tennis skills.

Acknowledgements

This work was financially supported in the part by the Kansai University Expenditures for Support of Establishing Research Centers, “Construction of Bridge Diagnosis Scheme by Brain Recognition Robotics” 2013. In addition, this work was financially supported in the part by Japan Construction Information Center Foundation (JACIC) Grant, 2015.

References

- Carpenter GA, Grossberg S, Reynolds J. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 1991; 4-5: 565-588
- Dietterich TG. Ensemble methods in machine learning. *The First International Workshop on Multiple Classifier System (MCS2000)*, 2000; 1-15
- Freund Y. Boosting a weak learning algorithm by majority. *Information and Computation*, 1995; 121; 2: 256-285
- Hayashi I. Acquisition of fuzzy rules using fuzzy ID3 with ability of learning for AND/OR operators. 1996 *Australian New Zealand Conference on Intelligent Information Systems (ANZIIS96)*, 1996; 187-190
- Hayashi I, Maeda T, Fujii M, Wang S, Tasaka T. Acquisition of embodied knowledge on sport skill using TAM network. 2009 *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE2009)*, 2009; 1038-1043
- Hayashi I, Maeda T, Ozawa J. A proposal of fuzzy ID3 with ability of tuning for AND connectives. *Japan Society for Fuzzy Theory and Systems*, 1999; 11(4): 677-682
- Hayashi I, Umamo M. Perspectives and trends of fuzzy neural networks. *J Fuzzy Theory and Systems*, 1993; 5(2): 178-190
- Hayashi I, Umamo M, Maeda T, Bastian A, Jain LC. Acquisition of fuzzy knowledge by NN and GA - A survey of the fusion and union methods proposed in Japan. *The Second International Conference on Knowledge-Based Intelligent Electronic Systems. KES'98 (Cat. No.98EX111)*, 1996; 69-78
- Hayashi I, Williamson JR. Acquisition of fuzzy knowledge from topographic mixture networks with attentional feedback. *The International Joint Conference on Neural Networks (IJCNN '01)*, 2001; 1386-1391
- Kasai J, Mori T, Yoshimura M, Ota A. A study of the forehand stroke of table tennis by three dimensional analyses using DTL method. *Waseda J Human Sciences*, 1994; 7(1): 119-127
- Kawato M, Gomi H. A computational model of four regions of the cerebellum based on feedback-error learning. *J Biological Cybernetics*, 1992; 68(2): 95-103
- Kosko B. *Neural networks and fuzzy systems: a dynamical approach to machine intelligence*. Prentice Hall

International Inc, Singapore; 1992

- Maeda T, Fujii M, Hayashi I, Tasaka T. Sport skill classification using time series motion picture data. *The 40th Annual Conference of the IEEE Industrial Electronics Society (IECON2014)*, 2014; 5272-5277
- Mochizuki Y, Himeno R, Omura K. Artificial skill and a new principle in sports. *System, Control and Information*, 2002; 46(8): 498-505
- Pandit SM, Wu SM. Time Series and System Analysis with Applications. *John Wiley & Sons*, 1983
- Parisi GI, von Stosch F, Magg S, Wermter S. Learning Human Motion Feedback with Neural Self-Organization. *2015 International Joint Conference on Neural Networks (IJCNN2015)*, 2015; 2973-2978
- Perl J, Baca A. Application of neural networks to analyze performance in sports. *The 8th Annual Congress of the European College of Sport Science*; 2003
- Quinlan JR. *C4.5: Programs for Machine Learning*. Burlington: Morgan Kaufmann Publishers; 1993
- Schapire RE. The Boosting approach to machine learning: An overview. In: *Nonlinear estimation and classification*. Denison, D.D, Hansen, M.H, Holmes, C, Mallick, B, and Yu, B, eds. Berlin, Germany, Springer; 2003
- Shiose T, Sawaragi T, Kawakami K, Katai O. Technological scheme of skill succession from ecological psychological approach. *J Japan Society for Ecological Psychology*, 2004; 1(1): 11-18
- Umano M, Okamoto H, Hatono I, Tamura H, Kawachi F, Umezu S, Kinoshita J. Fuzzy decision trees by fuzzy ID3 algorithm and its application to diagnosis systems. *The 3rd IEEE Conference on Fuzzy Systems*, 1994; 3; 2113-2118
- Williamson JR. Self-organization of topographic mixture networks using attentional feedback. *Neural Computing*, 2001; 13(3): 563-593

Corresponding author:

Isao Hayashi, Dr. Eng.

Graduate School of Informatics

Kansai University

2-1-1, Ryozenji-cho, Takatsuki, Osaka 569-1095 Japan

Phone: +81-72-690-2448

Fax: +81-72-690-2491

E-mail: ihaya@cpii.kutc.kansai-u.ac.jp