AVAILABLE AT
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Neural Networks 16 (2003) 1403–1420
Neural Networks
www.elsevier.com/locate/neunet

Adaptive categorization of ART networks in robot behavior learning using game-theoretic formulation

Wai-keung Fung*, Yun-hui Liu

Department of Automation and Computer-Aided Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China

Received 9 August 2000; revised 31 March 2003; accepted 31 March 2003

Abstract

Adaptive Resonance Theory (ART) networks are employed in robot behavior learning. Two of the difficulties in online robot behavior learning, namely, (1) exponential memory increases with time, (2) difficulty for operators to specify learning tasks accuracy and control learning attention before learning. In order to remedy the aforementioned difficulties, an adaptive categorization mechanism is introduced in ART networks for perceptual and action patterns categorization in this paper. A game-theoretic formulation of adaptive categorization for ART networks is proposed for vigilance parameter adaptation for category size control on the categories formed. The proposed vigilance parameter update rule can help improving categorization performance in the aspect of category number stability and solve the problem of selecting initial vigilance parameter prior to pattern categorization in traditional ART networks. Behavior learning using physical robot is conducted to demonstrate the effectiveness of the proposed adaptive categorization mechanism in ART networks.

Keywords: Adaptive categorization; Adaptive resonance theory; Vigilance parameter; Game theory; Nash equilibrium; Robot behavior learning; The BLOM architecture

1. Introduction

Robot behavior learning has been an emerging research topic for nearly two decades and fruitful results have been obtained through the effort of researchers. Various learning methodologies and techniques have been developed for robot learning (Connell & Mahadevan 1993; Franklin, Mitchell, & Thrun, 1996). However, existing results focus only on learning specific tasks on specific problem domains. Prior knowledge of the learning tasks, like robot dynamic and behavior models and the nature of the tasks that the robots are going to learn, is incorporated into the derivation of learning algorithms or the design of learning architectures. Some algorithms also impose restricted assumption on the learning tasks so as to simplify the learning processes and architectural design for robots. The developed learning algorithms and learning architectures are then task-dependent and problem-dependent. Our work, on the other hand, focuses on developing task-independent and problem-independent robot behavior learning techniques.

A robot behavior can be considered as a *sensorimotor mapping* from robot perceptual space to action space (Fung & Liu, 1998). The perceptual space is constructed from possible sensor data patterns while the action space is constructed from possible robot commands to drive robot motion. Since the perceptual space and action space are high-dimensional and continuous spaces, it is difficult to construct the whole sensorimotor map for each particular situation and action pair from finite number of training data patterns. In order to approximate this sensorimotor map, the perceptual and action spaces are divided into several categories (clusters) and mapping between the categories on the two spaces can be constructed. Within each category, situations (or actions) are similar in nature since it is assumed that similar situations (stimuli) invokes similar actions. The construction of sensorimotor map for particular behavior is then simplified from a high dimensional, nonlinear and discontinuous mapping to a set of simple *category mapping* (Fung & Liu, 1998). A generic neural network based architecture, called the Behavior Learning/Operating Modular (BLOM) Architecture, for

* Corresponding author. Tel.: +852-2609-8056; fax: +852-2603-6002.
E-mail addresses: wkfung@acae.cuhk.edu.hk (W.-k. Fung), yhliu@acae.cuhk.edu.hk (Y. Liu).

0893-6080/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0893-6080(03)00080-7

Nomenclature

Γ_{AC}	adaptive categorization game
$\mathscr{L}_{\mathrm{AC}}^{(i)}$	learning automaton associated with the ith F_2 neuron
Σ	state set $\{\mathbf{R}, \mathbf{r}, \mathbf{f}\}$
n	total number of F_2 neurons
k	number of F_2 neurons in RESET state
$\mathcal{I}_{(t)}$	index set of F_2 neurons in states RESONANCE and RESET after the t-th pattern presentation and $\mathcal{I}_j$ is the j-th element of $\mathcal{I}_{(t)}$
ρ_i	vigilance parameter of the ith F_2 neuron
ρ_i^*	updated vigilance parameter, $\rho^* = [\rho^*_{\mathcal{I}_1}, \ldots, \rho^*_{\mathcal{I}_{z_{k+1}}}]^{\mathrm{T}} \in \mathbb{R}^{k+1}$
μ_i	matching score for the ith F_2 neuron
$\beta_{\mathrm{VT}}^{(i)}$	cost coefficient of the cost function of attending Vigilance Test for the ith F_2 neuron and $\beta_{\mathrm{VT}} \triangleq [\beta_{\mathrm{VT}}^{(1)}, \beta_{\mathrm{VT}}^{(1)}, \ldots, \beta_{\mathrm{VT}}^{(k+1)}]^{\mathrm{T}} \in \mathbb{R}^{k+1}$
$\pi(\mathrm{i})$	payoff function of the ith F_2 neuron
$\pi_{\mathrm{R}}^{(i)}, \pi_{\mathrm{r}}^{(i)}, \pi_{\mathrm{f}}^{(i)}$	net gain of the ith F_2 neuron at states RESONANCE, RESET and FAIL, respectively
$\mathbf{\Psi}, \mathbf{\Xi}$	coefficient matrices in Nash equilibrium relationship between $(\rho^*, \beta_{\mathrm{VT}})$ pair
X	three-dimensional unit simplex
$\xi_{\mathrm{R}}^{(i)}, \xi_{\mathrm{r}}^{(i)}, \xi_{\mathrm{f}}^{(i)}$	state probabilities for the ith F_2 neuron in states RESONANCE, RESET and FAIL, respectively
$p_{uv}^{(i)}$	state transition probability of the ith F_2 neuron from state σ_u to σ_v, where $\sigma_u, \sigma_v \in \Sigma$
$q_{vu}^{(i)}$	confirmatory state transition probability of the ith F_2 neuron from state σ_u to σ_v, where $\sigma_u, \sigma_v \in \Sigma$
θ^i	reinforcement signal received by the ith F_2 neuron, $\theta^i \in \{-1, 0, 1\}$
a, b	learning rates for $\mathscr{L}_{\mathrm{AC}}^{(i)}$
$x_{\mathrm{R}}, x_{\mathrm{r}}, x_{\mathrm{f}}$	reward strengths for a F_2 neuron in states

robot behavior learning is employed to construct category mappings for behavior learning (Fung & Liu, 1998). The BLOM architecture consists of two sets of categorization networks in the perceptual and action domains (implemented by Fuzzy ART networks (Carpenter, Grossberg, & Rosen, 1991a) connected by a set of associative memories (implemented by Fuzzy Associative Memories FAM (Kosko, 1992). The categorization networks in the BLOM architecture are responsible for individual categorization of the input sensor data patterns (S-patterns) and action patterns (A-pattern). ART networks, thus, play an important role in robot behavior learning and enhancements (adaptive categorization) on ART networks benefit the behavior learning process.

Two of the difficulties arise in robot online behavior learning and they are described in the followings:

(1) When a robot is situated in an environment for operation, novel situations will usually be encountered by the robot when it interacts with its environment. The robot, which is capable of online learning, should update its knowledge-base with the novel situations as input stimuli. The robot knowledge-base should be updated incrementally and without overwriting any learned knowledge. The size of the knowledge-base will then expand with time as the robot starts to interact with the environment. The memory requirement of the robot knowledge-base during the learning process increases exponentially with time. This memory explosion phenomenon burdens the launch of incremental learning capability of the robot.

(2) In general, the robot perceptual space is a high dimensional, continuous, and space and the sensorimotor map constructed from behavior learning is nonlinear and discontinuous. It is impossible to construct the exact sensorimotor map from existing behavior learning methods. Approximation techniques are introduced in the mapping establishment, including partitioning or categorizing the input and output domains of the mapping. However, it is difficult for operators to specify the accuracy of behavior learning approximation and in a priori. Moreover, attention control in learning provides balanced utilization of resources (memory and computational effort) during learning. For instance, subspaces in a robot perceptual space can be categorized in a fine (coarse) scale when the robot pays much (less) attention into the particular area and vice versa.

In order to remedy the aforementioned difficulties in behavior learning, an adaptive categorization mechanism is developed for ART networks for perceptual and action patterns categorization by changing vigilance parameter ρ. ART networks refer to all neural networks developed based on the Adaptive Resonance Theory proposed by Grossberg and Carpenter in mid 1980s, including ART 1 (Carpenter & Grossberg, 1987a), ART 2 (Carpenter & Grossberg, 1987b), Fuzzy ART (Carpenter et al., 1991a), and their variants. This mechanism can be easily incorporated into all ART networks, including Fuzzy ART networks. The notion of adaptive categorization is first introduced to ART networks so that the granularity of categorization can be adjusted during learning for the adaptation to the dynamic environment of data patterns. Existing methods just blindly increase the vigilance parameter by a fixed amount when all committed F_2 neurons are exhausted (Vlajic & Card, 1998). This approach will eventually set vigilance parameter to 1 and as a result any new pattern will form its own cluster. In order to solve the problem, we propose a game-theoretic formulation on the adaptive vigilance parameter strategy in ART networks in this paper. We formulate the adaptive ART categorization mechanism as an adaptive categorization game Γ_{AC} in which the vigilance parameter is updated based on the *Nash Equilibrium* of Γ_{AC}. The granularity of categorization can be adjusted during learning for the adaptation to the dynamic environment of data patterns. To study the asymptotic behavior of the game Γ_{AC}, learning automata theory is also introduced in the repeated game analysis of the game Γ_{AC}. The proposed approach only leads to minor modifications (Section 3) to the original design of ART networks. Moreover, the modification method is independent of the fundamental categorization mechanism of ART networks. The game-theoretic vigilance adaptation strategy improves the clustering performance of ART networks in the aspect of category number stability, despite of the prespecified initial vigilance parameter is. Therefore, it is possible to avoid the problem of choosing suitable vigilance parameter in advance or data categorization by the trial-and-error approach. In other words, the proposed adaptive categorization mechanism for ART networks can remedy the two listed difficulties in behavior learning using the BLOM architecture.

This paper is organized as follows. Section 2 gives a brief introduction to a general architecture of ART networks. Section 3 describes the proposed game-theoretic formulation of adaptive categorization mechanism in ART networks and its properties. Moreover, Section 4 presents the ART networks based BLOM architecture for robot behavior learning and behavior learning experiments to show the effectiveness of adaptive categorization in behavior learning. In additions, Section 5 gives the conclusions of the paper.

2. ART networks

Since mid 1980s, Grossberg, Carpenter and their colleagues have proposed a series of ART networks based on the Adaptive Resonance Theory, which include ART-1 (Carpenter & Grossberg, 1987a) for binary inputs, ART-2 (Carpenter & Grossberg, 1987b) and Fuzzy ART (Carpenter et al., 1991a) for both binary and analog inputs, ARTMAP (Carpenter, Grossberg, & Reynolds, 1991b) and Fuzzy ARTMAP (Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992) for association between two data sets. ARTMAP and Fuzzy ARTMAP are supervised learning networks with two ART networks connected by a map field of neurons for associations establishment between categories formed from the two ART networks (Carpenter et al., 1991b).

Basically, ART networks form a class of self-organizing, self-stabilizing and self-scaling unsupervised competitive neural networks for categorization. ART networks solve the *Stability–Plasticity Dilemma*, which is also faced by other categorization and learning systems. The learned categorization codes are stable to resist the erosion of irrelevant data while being sensitive to novel data patterns. These features enable ART networks to be a powerful tool for incremental categorization learning, which is an important feature for on-line learning tasks. The granularity of the clusters (categories) is controlled by a fixed scalar parameter called vigilance parameter $\rho \in [0,1]$. The higher the vigilance parameter, the higher the granularity of the categories formed by ART networks.

In general, an ART network consists of two fully interconnected neuronal layers (as shown in Fig. 1 called F_1 and F_2, respectively. The F_1 layer is responsible for contrast enhancement and noise suppression on input data patterns. The number of F_1 neurons equals to the input pattern dimension m while the number of F_2 neurons reflects the capacity of categories (clusters) supported by the network n (Fig. 1). The F_2 layer forms a *winner-take-all* competitive layer. There are two sets of weights (long term memory (LTM)) connecting both the F_1 and F_2 layers, namely bottom-up weight ($F_1 \rightarrow F_2$) and top-down weight ($F_2 \rightarrow F_1$), respectively. The LTM represents the template feature vectors of the corresponding category. The neuronal activations are called short term memory (STM).

When a data pattern is passed to an ART network via F_1 layer, the activation pattern (STM) of F_1 neurons is gated by the bottom-up weights and is passed to F_2 layer. The F_2 neuron, that has the maximum matching score with the input pattern, is identified by the Matching Score test (MST). The activation pattern of the winning F_2 neuron (category), which is then gated by the top-down weights, is passed back to F_1 layer. This gated F_2 STM compares with the F_1 STM and the Vigilance Test (VT) follows. If they match with a level above a given vigilance level ρ, the network is said to be in *resonance* state and learning occurs by updating the winning (committed) cluster template vectors with the corresponding F_1 STM. If the match level is below

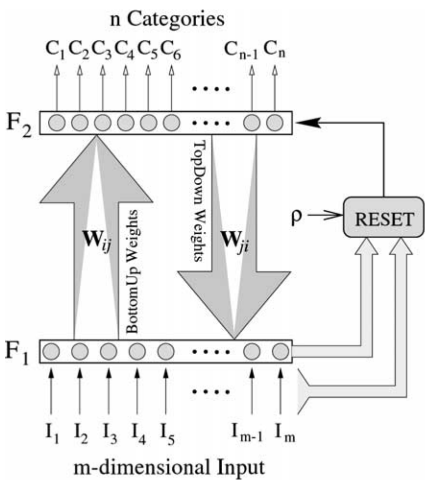


Fig. 1. Schmematics of ART networks.

the vigilance level, a *reset* signal is generated to the F_2 layer and the winning F_2 neuron is deactivated. The search for winning F_2 neurons continues until the match level is above the defined ρ and learning is conducted. The complicated reset mechanism in the two level competitive clustering is governed by '$\frac{2}{3}$ rule' conducted in the gain control nodes (as shown in Fig. 1) (Carpenter & Grossberg, 1987a).

3. Adaptive categorization in ART networks

According to the above discussion, the granularity of data categories is controlled by the vigilance parameter ρ. Original ART networks only use fixed vigilance parameter for all clusters and thus only fixed size clusters are formed. Fixed size clusters are difficult to represent thoroughly the data subspace and misclassification often happens in categorization-based classification.[1] On the other hand, variable size clusters, that are generated by adaptive vigilance parameter mechanism, have the capability of approximating data pattern subspace well and rendering decision boundaries, and thus misclassification can be avoided. Moreover, adaptive categorization helps preventing misclassification with data patterns from disjoint distributions. In additions, attention selectivity can be achieved by adaptive categorization so that fine clustering is applied on important data subspace while coarse clustering is applied on less important regions of the data pattern space. No mathematical formulation for ρ-adaptation has been proposed for ART networks in literatures. Few papers discussed varying vigilance parameters during categorization and they only suggest to *blindly* increase vigilance parameter by a fixed amount until all available F_2 neurons are exhausted (reset) for each data presentation (Vlajic & Card, 1998). One problem of this strategy is that each data pattern will be assigned to individual category eventually. In order to solve this problem, this paper proposes a vigilance parameter adaptation mechanism to adaptive categorization in ART networks.

We have made two modifications on classical ART networks so as to achieve adaptive categorization in ART networks, namely,

(1) The ART network assigns each F_2 neuron an individual vigilance parameter ρ_i, where $i = 1, 2, \ldots, n$, instead of assigning only one vigilance parameter for all F_2 neurons; and
(2) The individual vigilance parameter ρ_i for each F_2 neuron is adaptively changed based on the game-theoretic formulation on the competitive clustering mechanism in ART networks during categorization process.

[1] A categorization-based classifier is usually constructed by a categorization network, like ART networks, and a discrimination network that maps the categories formed into different classes for particular application.

3.1. Game-theoretic formulation

The competitive clustering mechanism in ART networks is formulated as an infinite n-person non-cooperative game. Game-theoretic analysis has been a popular technique in analyzing economic phenomena and strategies and human behaviors in complex systems. Game theory (Fudenberg & Tirole, 1991) is a mathematical technique for finding optimal (or sub-optimal) policy for individual agents with *conflicting* goals interacting in the same environment. Since most of the clustering algorithms, including ART networks, are competitive learning in nature, it is natural to employ game theory for analysis. The conflict exists in the ART adaptive categorization is that all F_2 neuron tries to have the presented data pattern categorized into the cluster it is representing. Vigilance parameter ρ adaptation strategy will be derived based on the *Nash Equilibrium* of the game-theoretic formulation of ART networks.

Each F_2 neuron is modeled as an individual player (decision maker) in the (adaptive) ART clustering process. An infinite non-cooperative n-persons game Γ_{AC} is defined as a triplet, $\Gamma_{AC} = (P, \{R^{(i)}\}_{i\in P}, \{\pi^{(i)}\}_{i\in P})$, where $P = \{1, 2, \ldots, n\}$, which is the index set of all players of the game Γ_{AC}, $R^{(i)}$ is the *strategy set* of the ith player and $\pi^{(i)}$ is the *payoff* function for the ith player. The vigilance parameter for each F_2 neuron forms its strategy $\rho_i \in R^{(i)}$ in ART networks and ρ_i is usually bounded in $[0, 1]$. $\pi^{(i)}$ is defined on the Cartesian product of the strategy sets of the players $R = \prod_{i\in P} R^{(i)}$ and $\pi^{(i)} : R \mapsto \mathbb{R}$.

As mentioned previously, each F_2 neuron must attend two independent tests for each pattern presentation: the *matching score test* (MST) and the *vigilance test* (VT). F_2 neurons can be classified into three groups when a data pattern is presented to an ART network according to three possible states:

Resonance state. Only one F_2 neuron is in the *resonance* state. This resonant F_2 neuron has passed both the matching score test and the vigilance test. The presented data pattern is assigned to the category represented by this particular F_2 neuron.
Reset state. Neurons in the reset state have passed the matching score test but failed the vigilance test. Denote the number of neurons is in the *reset* state by k.
Fail state. Neurons in the *fail* state have failed both the matching score test and the vigilance test. There are $(n - k - 1)$ F_2 neurons in this state.

In other words, each F_2 neuron is in one of the three possible states in the set

$$\Sigma = \{\mathrm{RESONANCE}\ (\mathbf{R}), \mathrm{RESET}(\mathbf{r}), \mathrm{FAIL}(\mathbf{f})\}$$

after every data presentation (Fig. 2). In our algorithm, only F_2 neurons, that are in states RESONANCE or RESET may have their vigilance parameters updated for next

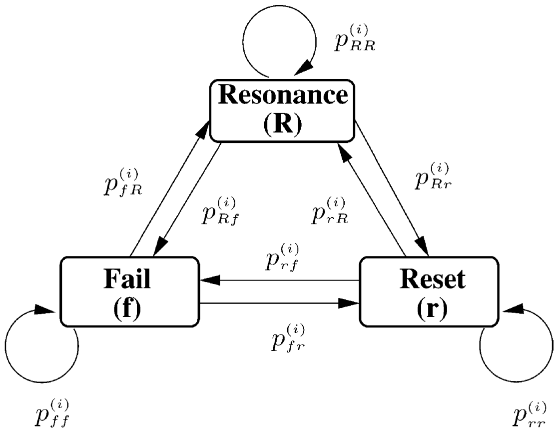


Fig. 2. State transitions of a F_2 neuron with each arc representing the state transition probability $p^{(i)}_{uv}$.

pattern presentation. The ρ-adaptation strategy then depends only on vigilance test mechanism. If a pattern is categorized with *direct access* (no F_2 neuron is in RESET state), no update on vigilance parameters will be conducted.

The basic derivation steps of adaptive categorization Γ_{AC} follows the Cournot game for oligopolic market model (Fudenberg & Tirole, 1991). The adaptive categorization scenario of ART networks is analogous to an oligopoly market, with the F_2 neurons corresponding to companies involved in the market and vigilance parameters of F_2 neurons corresponding to the prices of the products of the companies involved. Each F_2 neuron incurs *costs* when it attends the matching score test and the vigilance test and acquires *rewards* if it passes the tests. In the matching score test, the cost incurred and reward received by a F_2 neuron depend only on the matching score μ_i of that F_2 neuron. On the other hand, the cost incurred and reward received by a F_2 neuron depend only on the vigilance parameter ρ_i of that F_2 neuron in the vigilance test. The incurred costs of the matching score test $c^{(i)}_{\mathrm{MST}}$ and the vigilance test $c^{(i)}_{\mathrm{VT}}$ by the ith F_2 neuron form linear relations with μ_i and ρ_i, respectively, as follows,

$$c^{(i)}_{\mathrm{MST}} = \alpha_{\mathrm{MST}} + \beta_{\mathrm{MST}}\mu_i \tag{1}$$

$$c^{(i)}_{\mathrm{VT}} = \alpha_{\mathrm{VT}} - \beta^{(i)}_{\mathrm{VT}}\rho_i \tag{2}$$

where α_{MST}, β_{MST}, α_{VT} and $\beta^{(i)}_{\mathrm{VT}}$ are positive constants. The cost $c^{(i)}_{\mathrm{VT}}$ increases with decreasing ρ_i as the F_2 neurons are encouraged to have fine clustering (high ρ) for better approximation to data subspace. The rewards of the matching score test $r^{(i)}_{\mathrm{MST}}$ and the vigilance test $r^{(i)}_{\mathrm{VT}}$ obtained by the ith F_2 neuron, on the other hand, are given as

$$r^{(i)}_{\mathrm{MST}} = \mu_i\left((n-1)\mu_i - \sum_{j\neq i}\mu_j\right) \tag{3}$$

$$r^{(i)}_{\mathrm{VT}} = rho;_i\left(\sum_{\substack{j\in I(t)\\ j\neq i}}\rho_j - k\rho_i\right) \tag{4}$$

where n is the total number of F_2 neurons involved in categorization and k is the number of F_2 neurons in RESET state.

Each F_2 neuron will try its best to win the matching score test and tend to put as small effort (ρ_i) as possible to win the vigilance test. The reward $r^{(i)}_{\mathrm{VT}}$ and cost $c^{(i)}_{\mathrm{VT}}$ functions for the vigilance test demonstrate conflicting goals in the game-theoretic formulation. The net gain of F_2 neurons in the state $\sigma_i \in \Sigma$ are then given as follows,

$$\begin{aligned}
\pi^{(i)}_{\mathrm{R}} &= (r^{(i)}_{\mathrm{VT}} - c^{(i)}_{\mathrm{VT}}) + (r^{(i)}_{\mathrm{MST}} - c^{(i)}_{\mathrm{MST}})\\
\pi^{(i)}_{\mathrm{r}} &= -c^{(i)}_{\mathrm{VT}} + (r^{(i)}_{\mathrm{MST}} - c^{(i)}_{\mathrm{MST}})\\
\pi^{(i)}_{\mathrm{f}} &= -c^{(i)}_{\mathrm{MST}}
\end{aligned} \tag{5}$$

Denote the state of the ith F_2 neuron at the tth pattern presentation by $s^{(i)}(t)$ and let $I(t) \subset \mathrm{P}$ be the index set of F_2 neurons in states RESONANCE or RESET after the tth pattern presentation, that is $I(t) = \{i|s^{(i)}(t) = \mathbf{R}\} \cup \{i|s^{(i)}(t) = \mathbf{r}\}$. The payoff function $\pi^{(i)}$ of the ith ($i \in I(t)$) F_2 neuron at the tth pattern presentation is then defined as the expected gain of that F_2 neuron in the three possible states at the $(t+1)$-th pattern presentation,

$$\begin{aligned}
\pi^{(i)}(t) &= \mathrm{Prob}(s^{(i)}(t) = \mathbf{R})\pi^{(i)}_{\mathrm{R}}(t) + \mathrm{Prob}(s^{(i)}(t) = \mathbf{r})\pi^{(i)}_{\mathrm{r}}(t)\\
&\quad + \mathrm{Prob}(s^{(i)}(t) = \mathbf{f})\pi^{(i)}_{\mathrm{f}}(t)
\end{aligned} \tag{6}$$

where $\mathrm{Prob}(\cdot)$ denotes the probability of the given outcome.

3.2. State probability dynamics of F_2 neuron

A *learning automaton* $\mathscr{L}^{(i)}_{AC}$, $i = 1, 2, \ldots, n$, is constructed for each F_2 neuron to track the variations of state probabilities with time. A learning automaton (Narendra & Thathachar, 1989) is basically a *feedback* variable structure stochastic automaton system (Narendra & Thathachar, 1989) acting in an unknown stochastic environment so as to improve its performance under certain criteria. The learning automaton $\mathscr{L}^{(i)}_{AC}$ in the adaptive categorization game Γ_{AC} for each F_2 neuron consists of the following:

(1) a set of internal states $\Sigma = \{\mathbf{R}, \mathbf{r}, \mathbf{f}\}$;
(2) a set of output actions $\mathcal{O} = \{\mathbf{R}, \mathbf{r}, \mathbf{f}\}$;
(3) a set of input reinforcement signals $\theta^{(i)} \in \mathbf{\Theta} = \{-1, 0, 1\}$;
(4) a state transition probability matrix $\mathbf{P} \in \mathbb{R}^{3\times 3}$ that determines the state transition of $\mathscr{L}^{(i)}_{AC}$ at the next instant according to the current state. In other words, each element $p^{(i)}_{uv}$ in matrix $\mathbf{P}$ is defined as the conditional probability of $\mathscr{L}^{(i)}_{AC}$ in state σ_v at time $(t+1)$ (or the $(t+1)$-th pattern presentation) given

that it is in state σ_u at time t, $p_{uv}^{(i)} \triangleq \mathrm{Prob}(s^{(i)}(t+1) = \sigma_v | s^{(i)}(t) = \sigma_u)$;

(5) a reinforcement scheme $\Re$ for action probability update.

In our formulation of $\mathscr{L}_{AC}^{(i)}$, the action set $\mathcal{O}$ is the same as the state set Σ and each state can only have action that is the same as its state. Therefore, the corresponding output mapping $\mathcal{G}: \Sigma \rightarrow \mathcal{O}$ is an identity and state probability $\xi_u^{(i)}(t) \triangleq \mathrm{Prob}(s^{(i)}(t) = \sigma_u)$ is equivalent to the action probability. The state of a F_2 neuron is reflected in the state of learning automaton associated to it. The learning automaton associated for each F_2 neuron can then be characterized by a quadruple $\mathscr{L}_{AC}^{(i)} = \{\Sigma, \mathbf{\Theta}, \mathbf{P}, \Re\}$.

Fig. 3 depicts the interactions between the game Γ_{AC} or learning automaton and the environment. The environment is assumed to generate data patterns that the ART network will categorize. The environment is a so-called *Q-model environment* (Narendra & Thathachar, 1989) as it generates a finite set of discrete reinforcement signals (i.e. the set $\mathbf{\Theta}$ is finite) and $|\mathbf{\Theta}| > 2$. The purpose of reinforcement signal, which depends on the current state $s^{(i)}(t) \in \Sigma$ of $\mathscr{L}_{AC}^{(i)}$, is to guide the state probability adjustment according to the tracking capability of the learning automaton to its environment.

Reinforcement scheme $\Re$ of the learning automaton $\mathscr{L}_{AC}^{(i)}$ of the ith F_2 neuron provides the update rule for state probabilities. The updated state probability depends on the current state probability $\xi_u^{(i)}(t)$, current state $s^{(i)}(t)$ and reinforcement signal $\theta^{(i)}(t)$ received. By the conditional probability theory, the state probability at the next instant $\xi_v^{(i)}(t+1) \triangleq \mathrm{Prob}(s^{(i)}(t+1) = \sigma_v)$ is given as

$$\xi_v^{(i)}(t+1) = \sum_{\sigma_u \in \Sigma} \{\mathrm{Prob}(s^{(i)}(t+1) = \sigma_v | s^{(i)}(t) = \sigma_u) \times \mathrm{Prob}(s^{(i)}(t) = \sigma_u)\} = \sum_{\sigma_u \in \Sigma} p_{uv}^{(i)} \xi_u^{(i)}(t) \tag{7}$$

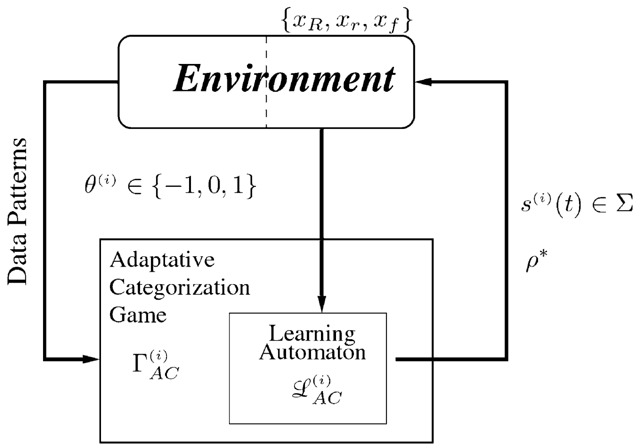


Fig. 3. Feedback connection of Adaptive Categorization Game $\Gamma_{AC}^{(i)}$ of the ith F_2 neuron.

Before describing the reinforcement scheme for the learning automata, we define the following concepts. The state space of state probability in $\mathscr{L}_{AC}$ is a three-dimensional unit simplex and is defined as $X = \{\xi | \sum_{\sigma_u \in \Sigma} \xi_u = 1;\ 0 \le \xi_u \le 1; \sigma_u \in \Sigma\}$ and the *confirmatory* transition probability $q_{vu}^{(i)}$, which is defined as

$$q_{vu}^{(i)} \triangleq \mathrm{Prob}(s^{(i)}(t) = \sigma_u | s^{(i)}(t+1) = \sigma_v) = p_{uv}^{(i)} \frac{\mathrm{Prob}(s^{(i)}(t) = \sigma_u)}{\mathrm{Prob}(s^{(i)}(t+1) = \sigma_v)} = p_{uv}^{(i)} \frac{\xi_u^{(i)}(t)}{\sum_{\sigma_u \in \Sigma} p_{uv}^{(i)} \xi_u^{(i)}(t)} \tag{8}$$

from the Bayes Theorem and Eq. (7). The reinforcement signal $\theta^{(i)}$, reflecting the tracking performance of the learning automaton on its situated environment, is defined based on the state transition probabilities $p_{uv}^{(i)}$ and $q_{vu}^{(i)}$,

$$\theta^{(i)} = \begin{cases} -1, & \text{if } s^{(i)} = \operatorname{argmin}_{\sigma_u \in \Sigma} \sum_{\sigma_v \in \Sigma} p_{uv}^{(i)} q_{vu}^{(i)} \\ 1, & \text{if } s^{(i)} = \operatorname{argmax}_{\sigma_u \in \Sigma} \sum_{\sigma_v \in \Sigma} p_{uv}^{(i)} q_{vu}^{(i)} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

The sum-and-product term $\sum_{\sigma_v \in \Sigma} p_{uv}^{(i)} q_{vu}^{(i)}$ measures the amount of evidence that supports the transition from state σ_u at the tth pattern presentation to state σ_v at the $(t+1)$-th pattern presentation. If the value of this sum-and-product is high (low), we believe that the learning automaton $\mathscr{L}_{AC}$ can (cannot) somehow predict the state transition of the corresponding F_2 neuron and $\mathscr{L}_{AC}^{(i)}$ is reinforced with $\theta = 1$ (punished with $\theta = -1$) and vice versa.

The general reinforcement scheme for state probabilities is given as follows,

$$\xi_u^{(i)}(t+1) = \begin{cases} \xi_u^{(i)}(t) - \frac{1}{2}(1+\theta^{(i)}) g_u(\xi^{(i)}) \\ \quad + \frac{1}{2}(1-\theta^{(i)}) h_u(\xi^{(i)}), & \text{if } s^{(i)}(t) \neq \sigma_u \\ \xi_u^{(i)}(t) + \frac{1}{2}(1+\theta^{(i)}) \sum_{\sigma_v \neq \sigma_u} g_v(\xi^{(i)}) \\ \quad - \frac{1}{2}(1-\theta^{(i)}) \sum_{\sigma_v \neq \sigma_u} h_v(\xi^{(i)}) & \text{if } s^{(i)}(t) = \sigma_u \end{cases} \tag{10}$$

where $g_u(\xi^{(i)})$ and $h_u(\xi^{(i)})$ are the reward and penalty functions, respectively, and they are nonnegative and continuous functions $X \mapsto (0,1)$ for $\sigma_u \in \Sigma$ (Narendra & Thathachar, 1989). In order to guarantee that $\xi^{(i)}(t+1) \in X$, we have

$$\begin{cases} 0 < g_u(\xi^{(i)}) < \xi_u^{(i)}, \\ 0 < \sum_{\sigma_v \neq \sigma_u} (h_v(\xi^{(i)}) + \xi_v^{(i)}) < 1, \end{cases} \quad \forall \sigma_u \in \Sigma \tag{11}$$

(Narendra & Thathachar, 1989). One of the common linear reward and penalty functions pair, which satisfies all the aforementioned requirements in Eq. (11), is given as with scalar learning rates $0<a<1$ and $0<b<1$,

$$g_u(\xi^{(i)}(t))=a\xi_u^{(i)}(t)$$

$$h_u(\xi^{(i)}(t))=b(\tfrac{1}{2}-\xi_u^{(i)}(t))$$

(Narendra & Thathachar, 1989). Then the reinforcement scheme $\Re$ for learning automaton $\mathscr{L}_{AC}^{(i)}$ becomes

$$\xi_u^{(i)}(t+1) = \begin{cases} \xi_u^{(i)}(t)-\frac{1}{2}a(1+\theta^{(i)})\xi_u^{(i)}(t) & \\ +\frac{1}{2}b(1-\theta^{(i)})(\frac{1}{2}-\xi_u^{(i)}(t)), & \text{if } s^{(i)}(t)\neq\sigma_u \\ \xi_u^{(i)}(t)+\frac{1}{2}a(1+\theta^{(i)})(1-\xi_u^{(i)}(t) & \\ -\frac{1}{2}b(1-\theta^{(i)})\xi_u^{(i)}(t) & \text{if } s^{(i)}(t)=\sigma_u \end{cases} \quad (12)$$

3.3. ρ Adaptation—Nash equilibrium of Γ_{AC}

The payoff function $\pi^{(i)}(t)$ of the i-th ($i\in\mathcal{I}(t)$ F_2 neuron can be derived from Eqs. (1)–(6) and is given as

$$\pi^{(i)}(t)=\xi_{\mathrm{R}}^{(i)}\left(\sum_{j\neq i}\rho_j-k\rho_i\right)\rho_i+(\xi_{\mathrm{R}}^{(i)}+\xi_{\mathrm{r}}^{(i)})\beta_{\mathrm{VT}}^{(i)}\rho_i + \mathcal{A}(\alpha_{\mathrm{VT}},\alpha_{\mathrm{MST}},\beta_{\mathrm{MST}},\xi_{\mathrm{R}}^{(i)},\xi_{\mathrm{r}}^{(i)},\xi_{\mathrm{f}}^{(i)}) \quad (13)$$

where $\mathcal{A}$ is an expression that is independent of ρ_i. The Nash equilibrium of the adaptive categorization game Γ_{AC} can be deduced easily from the *best response function* of each player (F_2 neuron). The best response function of the ith F_2 neuron gives the best reply to strategies $\bar{\rho}_i=\rho\backslash\rho_i$ of other F_2 neurons (Fudenberg & Tirole, 1991). The best response function of a F_2 neuron gives its best expected payoff according to the strategies (vigilance parameters) of the other F_2 neurons. The best response function of the ith F_2 neuron is then given by setting $\partial\pi^{(i)}/\partial\rho_i=0$, where $i\in\mathcal{I}(t)$,

$$\frac{\partial\pi^{(i)}}{\partial\rho_i}=-2k\xi_{\mathrm{R}}^{(i)}\rho_i+\xi_{\mathrm{R}}^{(i)}\sum_{j\neq i}\rho_j+\beta_{\mathrm{VT}}^{(i)}(\xi_{\mathrm{R}}^{(i)}+\xi_{\mathrm{r}}^{(i)}) \quad (14)$$

Setting $\partial\pi^{(i)}/\partial\rho_i=0$, we have a linear equation system of $|\mathcal{I}(t)|=k+1$ equations,

$$2k\rho_i^*-\sum_{j\neq i}\rho_j^*=\beta_{\mathrm{VT}}^{(i)}\left(1+\frac{\xi_{\mathrm{r}}^{(i)}}{\xi_{\mathrm{R}}^{(i)}}\right),\qquad j\in\mathcal{I}(t) \quad (15)$$

The Nash equilibrium ρ^* of Γ_{AC} is defined as a strategy that satisfies the best response functions of all players so that strategies of the i-th F_2 neuron ($i\in\mathcal{I}(t)$) are the best replies to the strategies $\bar{\rho}_i$ of each other players. In other words, Nash equilibria occur where the best response functions of players cross. Therefore, the Nash equilibria of Γ_{AC} are given as pairs of $(\boldsymbol{\rho}^*,\boldsymbol{\beta}_{\mathrm{VT}})$ that satisfies the equation

$$\boldsymbol{\Psi}_\rho^*=\boldsymbol{\Xi}\beta_{\mathrm{VT}},$$

$$\underbrace{\begin{bmatrix} 2k & -1 & \cdots & -1 \\ -1 & 2k & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & 2k \end{bmatrix}}_{\boldsymbol{\Psi}\in\mathbb{R}^{(k+1)\times(k+1)}} \underbrace{\begin{bmatrix} \rho_{I_1}^* \\ \rho_{I_2}^* \\ \vdots \\ \rho_{I_{k+1}}^* \end{bmatrix}}_{\rho^*\in\mathbb{R}^{k+1}} = \underbrace{\begin{bmatrix} 1+\frac{\xi_{\mathrm{r}}^{(I_1)}}{\xi_{\mathrm{R}}^{(I_1)}} & & 0 \\ & \ddots & \\ 0 & & 1+\frac{\xi_{\mathrm{r}}^{(I_{k+1})}}{\xi_{\mathrm{R}}^{(I_{k+1})}} \end{bmatrix}}_{\boldsymbol{\Xi}\in\mathbb{R}^{(k+1)\times(k+1)}} \underbrace{\begin{bmatrix} \beta_{\mathrm{VT}}^{(I_1)} \\ \beta_{\mathrm{VT}}^{(I_2)} \\ \vdots \\ \beta_{\mathrm{VT}}^{(I_{k+1})} \end{bmatrix}}_{\boldsymbol{\beta}_{\mathrm{VT}}\in\mathbb{R}^{k+1}} \quad (16)$$

where $\mathcal{I}_j$ is the j-th element in the index set $\mathcal{I}(t)$.

At the Nash equilibrium of the game Γ_{AC}, the relationship between the $(\boldsymbol{\rho}^*,\boldsymbol{\beta}_{\mathrm{VT}})$ pair is given as

$$\begin{cases} \beta_{\mathrm{VT}}^{(i)}=\dfrac{\xi_{\mathrm{R}}^{(i)}}{\xi_{\mathrm{R}}^{(i)}+\xi_{\mathrm{r}}^{(i)}}\left(2k\rho_i^*-\displaystyle\sum_{j\neq i}\rho_j^*\right) \\ \rho_i^*=\dfrac{1}{k(2k+1)}\left((k+1)\left(1+\dfrac{\xi_{\mathrm{r}}^{(i)}}{\xi_{\mathrm{R}}^{(i)}}\right)\beta_{\mathrm{VT}}^{(i)} \quad i,j\in\mathcal{I}(t)\right. \\ \left.+\displaystyle\sum_{j\neq i}\left(1+\dfrac{\xi_{\mathrm{r}}^{(j)}}{\xi_{\mathrm{R}}^{(j)}}\right)\beta_{\mathrm{VT}}^{(j)}\right) \end{cases} \quad (17)$$

By substituting Eq. (17) into Eq. (13), the payoff $\pi_{\mathrm{NE}}(t)$ of the i-th F_2 neuron at Nash equilibrium is given as

$$\pi_{\mathrm{NE}}^{(i)}=k\xi_{\mathrm{R}}^{(i)}(\rho_i^*)^2+\mathcal{A}(\alpha_{\mathrm{VT}},\alpha_{\mathrm{MST}},\beta_{\mathrm{MST}},\xi_{\mathrm{R}}^{(i)},\xi_{\mathrm{r}}^{(i)},\xi_{\mathrm{f}}^{(i)}) \quad (18)$$

The one-to-one correspondence between the vectors ρ^* and β_{VT} at the Nash equilibrium of the game Γ_{AC} is obtained.

Every F_2 neuron eagers to gain as much payoff $\pi_{\mathrm{NE}}^{(i)}$ as possible in the competition for being in the RESONANCE state by tuning its ρ_i^* to 1 during categorization. However, it is not economical because the total energy supplied by all F_2 neurons in categorization process is not minimized so that all F_2 will eventually change their vigilance parameters to the extreme values.[2] Vigilance parameters are adapted so that minimum energy is consumed by the F_2 neurons so as to overcome the potential barrier in becoming the winning F_2 neuron during the categorization of data patterns.

The potential barrier $\mathbb{P}_i$ of avoiding the ith F_2 neuron from becoming a winning neuron (*ie.* in RESONANCE state) is defined as

$$\mathbb{P}_i\triangleq(1-\xi_{\mathrm{R}}^{(i)})\rho_i^* \quad (19)$$

Intuitively, the potential barrier increases with increasing vigilance parameter and the state probability $\xi_{\mathrm{R}}^{(i)}$ indicates the degree of easiness for the F_2 neuron to overcome

[2] This argument is analogous to the 'principle of least action' hypothesis proposed by Pierre–Louis Moreau de Maupertuis (1698–1759) in the field of analytical dynamics (Williams, 1996).

the potential barrier. $\xi_{\mathrm{R}}^{(i)}$ introduces an inhibitory effect on the F_2 neuron RESONANCE state potential barrier.

On the other hand, the 'kinetic' energy of the ith F_2 neuron measures the capability of the i-th F_2 neuron to have the current data pattern being categorized into the ith cluster. The 'kinetic' energy of the ith neuron can then be reflected by the payoff gained $\pi_{\mathrm{NE}}^{(i)}$ of the ith F_2 neuron at the Nash equilibrium of the adaptive categorization game Γ_{AC} (with vigilance parameter ρ_i^*) in the data pattern competition. In other words, the more 'kinetic' energy the F_2 neuron releases in competition, the more the payoff it will gain, and vice versa. Therefore, the 'kinetic' energy of the ith F_2 neuron is defined as $K_i \triangleq k\xi_{\mathrm{R}}^{(i)}(\rho_i^*)^2 + \mathcal{A}_i$, where $\mathcal{A}_i$ is the shorthand for the term that is independent of ρ_i^* in the expression for F_2 neuron payoff gain $\pi_{\mathrm{NE}}^{(i)}$ at Nash equilibrium. The difference between $\mathbb{K}_i$ and $\mathbb{P}_i$ is minimized subject to vigilance parameters of F_2 neurons, ρ_i^* $i \in \mathcal{I}(t)$, so that the F_2 neurons can consume the minimal energy to overcome the potential barriers in the next pattern presentation.

By defining the Lagrangian $\mathbb{L}_i$ for each F_2 neuron, $i \in \mathcal{I}(t)$, as

$$\mathbb{L}_i \triangleq \mathbb{K}_i - \mathbb{P}_i = k\xi_{\mathrm{R}}^{(i)}(\rho_i^*)^2 - (1 - \xi_{\mathrm{R}}^{(i)})\rho_i^* + \mathcal{A}_i \tag{20}$$

the updated vigilance parameter, ρ_i^* is given by setting $(\partial\mathbb{L}_i/\partial\rho_i^*) = 0$. Then the vigilance parameter for the i-th F_2 neuron at the tth pattern presentation is updated as

$$\rho_i^*(t) = \begin{cases} \dfrac{1 - \xi_{\mathrm{R}}^{(i)}}{2k\xi_{\mathrm{R}}^{(i)}}, & \text{if } \xi_{\mathrm{R}}^{(i)} > \dfrac{1}{2k+1} \\ \rho_i^*(t-1), & \text{otherwise} \end{cases} \tag{21}$$

where $i \in \mathcal{I}(t)$. The condition imposed in the vigilance parameters update law is to restrict each ρ_i^* to lie in its nominal range $(0, 1)$.

Algorithm 1. Adaptive categorization Γ_{AC} for ART networks.

1. *Initialization for traditional ART networks* (Carpenter & Grossberg, 1987a; Carpenter et al., 1991a),
2. Arbitrary set vigilance parameters ρ_i, learning rate a and b for learning automata $\mathcal{L}_{AC}^{(i)}$ for each F_2 neuron.
3. State probabilities initialization for each F_2 neuron, $\xi_u \leftarrow \frac{1}{3}$, $\sigma_u \in \{\mathbf{R}, \mathbf{r}, \mathbf{f}\}$
4. State transition probabilities initialization for each F_2 neurons $p_{uv}^{(i)} \leftarrow \frac{1}{3}$, σ_u, $\sigma_v \in \{\mathbf{R}, \mathbf{r}, \mathbf{f}\}$
5. $t \leftarrow 0$
6. **while** there is data pattern being fed into the network **do**
   7. *Perform the traditional categorization procedure of ART networks* (Carpenter & Grossberg, 1987a),
   8. Compute the confirmatory state transition probabilities $q_{vu}^{(i)}$ (Eq. (8))
   9. Assign reinforcement signal $\theta^{(i)}$ for each F_2 neuron according to Eq. (9).
   10. Update state probabilities $\xi_{u^{(i)}}$, $\sigma_u \in \Sigma$ based on the reinforcement scheme $\Re$ see Eq. (12)
   11. Update state transition probabilities $p_{uv}^{((i))}$ based on Bayes Theorem.
   12. Update vigilance parameters by Eq. (21).
   13. $t \leftarrow t+1$
14. **end while**

3.4. Repeated game analysis on the game γ_{AC}

The performance of ART networks with the game-theoretic adaptive categorization algorithm as $t \to \infty$ is investigated. The analysis focuses on asymptotic behavior of the game Γ_{AC} operating in a *stationary* environment. Since the dynamics of the game Γ_{AC} is governed by the learning automaton $\mathcal{L}_{AC}$, we study the asymptotic behavior of Γ_{AC} for each F_2 neuron through analyzing the behavior of the associated learning automata $\mathcal{L}_{AC}^{(i)}$ $i = 1, 2, \ldots, n$ with reinforcement scheme $\Re$ given in Eq. (12). Theoretically, a variable structure learning automaton in a stationary environment can be considered as a discrete Markov process $\{\xi(t)\}_{t\geq 0} \in X$ (White, 1993) with dynamics described by the reinforcement scheme in Eq. (12). Only characteristics of the game Γ_{AC} is listed in the followings and proofs of these properties are listed in Appendix A.

(1) The proposed reinforcement scheme (12) or the Markov process $\{\xi(t)\}_{t>0}$ is *strictly distance diminishing*.

(2) The Markov process $\{\xi(t)\}_{t>0}$ has absorbing state with probability 1 if and only if $b = 0$. This implies that F_2 neurons will stay in one of the states in Σ asymptotically and this leads to fixed categorization whatever data patterns are passed to the ART networks in a long run. Hence, all data patterns are categorized to a fixed category (F_2 neuron) asymptotically if $b = 0$. The probability of converging to a particular state (convergence probability) and rate of convergence of each F_2 neuron depends on its initial state probability $\xi(0)$ (Narendra & Thathachar, 1989). This describes the so-called 'reputation effect' of the adaptive categorization game Γ_{AC} when the game is played repeatedly (Fudenberg & Tirole, 1991).

(3) The Markov process $\{\xi(t)\}_{t>0}$ is *ergodic* and has a limiting expected state probability $\xi \triangleq \lim_{t\to\infty} \mathbf{E}[\xi(t)]$ if and only if $b \neq 0$. The asymptotic state probability depends only on the unknown environment characteristics called reward strength, $x_u \in (-1, 1)$ for each state $\sigma_u \in \Sigma$, which is defined as $x_u \triangleq \mathbf{E}[\theta(t)|s(t) = \sigma_u]$. If the reward strength x_u for state σ_u is larger than that of any other state σ_v, then state σ_u is chosen asymptotically with a higher probability than any other states σ_v on average.

(4) It is worth to be noted that the learning automaton $\mathscr{L}_{AC}$ demonstrates bifurcation phenomenon on b based on the previous two properties. The learning automaton $\mathscr{L}_{AC}$ exhibits distinct system behaviors in the cases of $b = 0$ and $b \neq 0$.

(5) If $b \neq 0$, the learning automaton $\mathscr{L}_{AC}$ is *expedient*, i.e. $\lim_{t\to\infty} \mathbf{E}[M(t)] > M_0$, where $M(t) \triangleq \mathbf{E}[\theta(t)|\xi(t)] = \sum_{\sigma_u \in \Sigma} x_u \xi_u(t)$ and $M_0 = (x_R + x_r + x_f)/3$ is the *norm of behavior* for pure-chance automaton (Narendra & Thathachar, 1989). The learning automaton $\mathscr{L}_{AC}$ is *absolutely expedient* if and only if $b = 0$. It is also ϵ-optimal in all stationary environment when $b = 0$ as absolutely expediency implies ϵ-optimality under all stationary environment (Narendra & Thathachar, 1989). This implies that the learning automaton $\mathscr{L}_{AC}$ can learn as it outperforms its pure-chance automaton counterpart) as it is expedient (even absolutely expedient when $b = 0$) for all the possible value of b.

(6) To sum up, the rule of thumb of parameter selection in reinforcement scheme derivation is that $0 < a < 1, 0 < b < 1$ and small b so that the adaptive categorization game Γ_{AC} can track the changing environment (when $b \neq 0$) while being relatively stable with limiting state (when $b = 0$). The limiting state indicates what it has learned about the characteristics of the environment. Moreover, the expected vigilance parameters of the game Γ_{AC} will also converge as the reinforcement scheme is ergodic and the formation of the updated vigilance parameters involves state probabilities (Eq. (16)).

3.5. Simulations

Simulations results are presented to compare the performances of Fuzzy ART networks (Carpenter et al., 1991a) with and without using the proposed game-theoretic ρ adaptation. Fuzzy ART network is selected for simulations because it is simple and it provides wide input pattern diversity (both binary and analog data patterns). Learning rates in the reinforcement scheme are set as $a = 0.75$ and $b = 0.1$.

Two simulations with different data patterns distributions are presented as follows,

(1) Two thousand uniformly distributed random two-dimensional data patterns, which are confined in a 'ring-like' region centered at $(0.45, 0.52)$ and with inner and outer radii of 0.15 and 0.405, respectively, are generated for simulations. Tests on Fuzzy ART networks with and without ρ-adaptation are performed with starting vigilance parameters at 0.4, 0.55, 0.7 and 0.85. The Fuzzy ART learning rates employed in all tests are set to 0.9 and the pattern order in all tests are the same. Fig. 4 depicts the categorization results in the tests. The number of categories formed in the tests are listed as in Table 1,

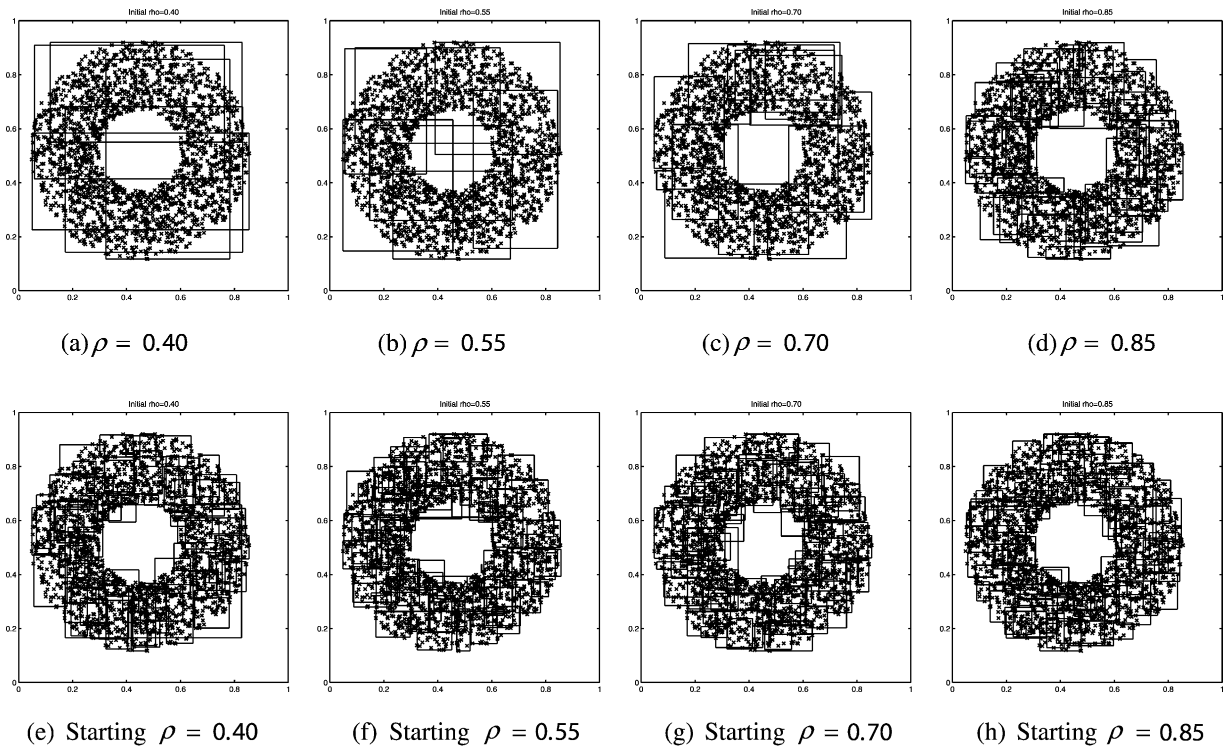


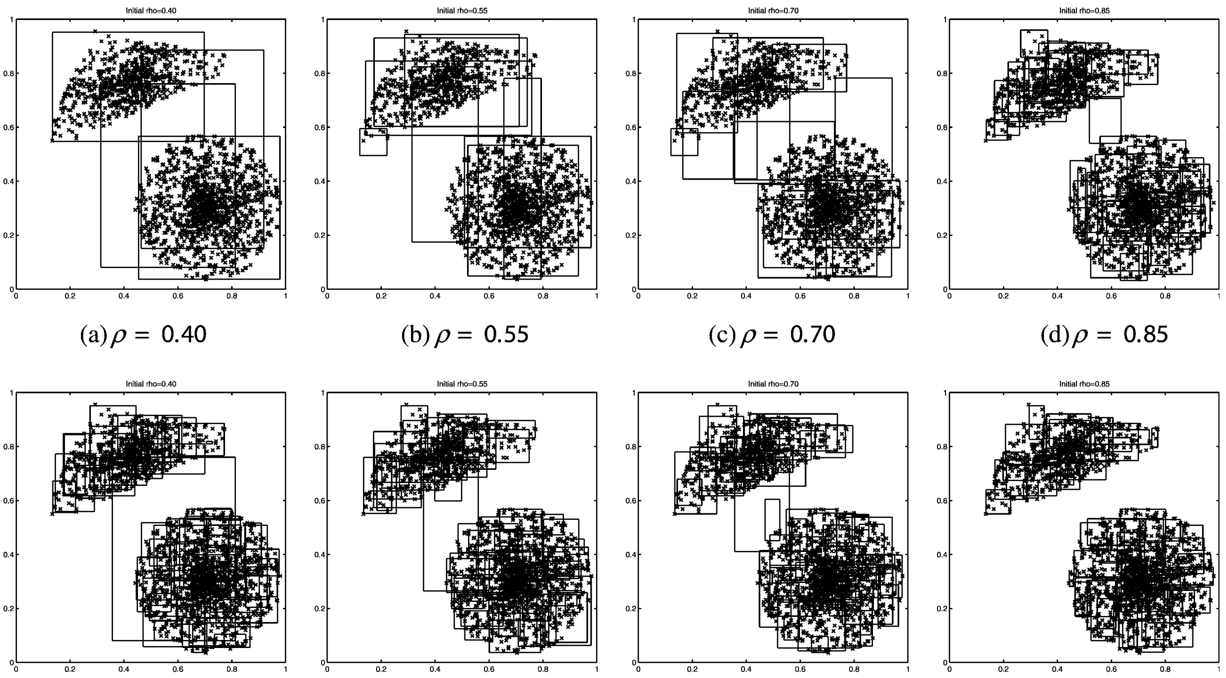
Fig. 4. Categorization of uniformly distributed data patterns in a 'ring-like' region by Fuzzy ART networks with (lower row) and without (upper row) ρ-adaptation.

Table 1
Categories number formed in all tests on a set uniformly distributed 2D data patterns in a 'ring-like' region

	Starting ρ			
	0.40	0.55	0.70	0.85
Non-adaptive ρ	5	7	13	47
Adaptive ρ	83	99	101	107
Final ρ	0.882	0.883	0.886	0.886

(2) Two thousand uniformly distributed random two-dimensional data patterns, which are confined in a pair of *disjoint* distributions, are generated for simulations. The data patterns are confined either in a circular region centered at $(0.710, 0.305)$ with radius 0.27 or a triangular regions with vertices at $(0.08, 0.49)$, $(0.28, 0.96)$ and $(0.87, 0.85)$, respectively. Tests on Fuzzy ART networks with and without ρ-adaptation are performed with starting vigilance parameters at 0.4, 0.55, 0.7 and 0.85. The Fuzzy ART learning rates employed in all tests are set to 0.9 and the pattern order in all tests are the same. Fig. 5 depicts the categorization results in the tests. The number of categories formed in the tests are listed as in Table 2.

As shown in Figs. 4 and 5, the Fuzzy ART prototypes (Carpenter et al., 1991a) generated are displayed as rectangles. The prototype rectangles generated by Fuzzy ART without ρ-adaptation render the pattern distribution boundary poorly, especially for lower fixed vigilance parameters. On the other hand, the prototype rectangles generated by Fuzzy ART with ρ-adaptation render the pattern distribution boundary well no matter what value the starting vigilance parameter is. Moreover, the number of categories generated by Fuzzy ART without ρ-adaptation grows geometrically with increasing starting vigilance parameters while the number of categories generated by Fuzzy ART with ρ-adaptation is much insensitive to the starting vigilance parameter chosen. Thus ρ-adaptation remedies the difficulties in choosing vigilance parameters prior to the data clustering process using ART networks. The updated vigilance parameters converge to similar values with different starting vigilance parameters after the 2000 data patterns categorization. In additions, as demonstrated in Simulation 2 (with patterns generated from disjoint distributions), the categories generated by Fuzzy ART with ρ-adaptation cover far less patterns from either of the disjoint distributions than that generated by conventional Fuzzy ART network, as shown in Fig. 5. The categories generated by ρ-adaptive Fuzzy ART network can even be divided into two distinct groups that contain patterns from one and only one distribution when the starting



(a) $\rho = 0.40$ (b) $\rho = 0.55$ (c) $\rho = 0.70$ (d) $\rho = 0.85$

(e) Starting $\rho = 0.40$ (f) Starting $\rho = 0.55$ (g) Starting $\rho = 0.70$ (h) Starting $\rho = 0.85$

Fig. 5. Categorization of uniformly distributed data patterns in a pair of disjoint triangular and circular regions by Fuzzy ART networks with (lower row) and without (upper row) ρ-adaptation.

Table 2
Categories number formed in all tests on a set uniformly distributed 2D data patterns in a pair of disjoint regions

	Starting ρ			
	0.40	0.55	0.70	0.85
Non-adaptive ρ	4	8	14	43
Adaptive ρ	74	71	73	73
Final ρ	0.874	0.874	0.874	0.874

vigilance parameter is 0.85. This helps to avoid misclassification in classifier systems that are constructed from Fuzzy ART networks.

4. Robot behavior learning

A robot behavior can be considered as a mapping from the perceptual space or sensor data space $\mathcal{S}$ to the action space $\mathcal{A}$, $\mathcal{E}: \mathcal{S} \to \mathcal{A}$, called a *sensorimotor map* (Fung & Liu, 1998). Sensor data space or perceptual space $\mathcal{S}$ is constructed from all sensors modalities installed on the robot while action space $\mathcal{A}$ is constructed from robot actuators output that are under control. The goal of behavior learning is then to construct the aforementioned sensorimotor maps for various robot behaviors. All inputs for this learning task are in a form of data patterns consisting of consecutive samples from sensors and actuators equipped on robots. Typical sensors involved in behavior learning include ultrasonic range sensors, tactile sensors, vision sensors and so on.

Since the sensor domain $\mathcal{S}$ and action domain $\mathcal{A}$ are high-dimensional and continuous spaces, it is difficult to construct the whole sensorimotor map for each particular situation and action pair from finite sets of training data patterns. In order to approximate this sensorimotor map, the sensor domain $\mathcal{S}$ and action domain $\mathcal{A}$ are divided into several categories (clusters) and mapping between the categories on the two domains can be constructed. Within each category, situations (or actions) are similar in nature. It is assumed that similar situations (stimuli) invokes similar actions. This simplifies the construction of sensorimotor map for particular behavior from a high dimensional, nonlinear and discontinuous mapping to a set of simple *category mappings*, from a particular situation category $\sigma \tilde{\subset} \mathcal{S}$ to action category $\alpha \tilde{\subset} \mathcal{A}$[3] (Fung & Liu, 1998). Biological evidenceof the categorization of situation and action spaces can be found inanimals (for example, African Grey Parrot (Perrberg, 1996) and humans (Massaro, 1990). This indicates that generalization on stimuli (situations) occurs within category while discrimination occurs among categories. Psychologists have been conducting numerous experiments to indicate the presence of stimulus generalization and are recorded in various psychological literatures on behavior and learning (Walker, 1995).

4.1. The BLOM architecture

Based on the sensorimotor map model of a behavior, a generic neuralnetwork based architecture for robot behavior learning, which iscalled the Behavior Learning/Operating Modular (BLOM) Architecture, isemployed to reconstruct an individual robot behavior (Fung & Liu, 1998). This architecture incorporates both the learning and operating modesin the same structure. The advantage of incorporating the learning and operating modes in the same structure is that the effort and loss of information in the transformation between the representations of knowledge in individual learning and operating modules are saved. The BLOM architecture, which is shown in Fig. 6, consists of two groups of categorization networks (implemented by Fuzzy ARTnetworks (Carpenter et al., 1991a)) connected by a set of associative memories (implemented by fuzzy associative memories FAM (Kosko, 1992)). The perceptual and action categorization networks, which are denoted as $\mathcal{CN}_S$ and $\mathcal{CN}_A$, respectively, are responsible for categorization of input sensor data patterns (S-patterns) and action patterns (A-pattern) individually. Each categorization network is assigned to categorize one dimension of data patterns. There is also a *coding layer*, which is made of a categorization network $\mathcal{CN}_C$, for perceptual code (S-code) compactification and distribution. The associative memory are for association establishments between perceptual and action categories. Action memories (implemented by OLAM (Kohonen, 1989)) are employed to store prototypical action patterns for each action category in order to reconstruct appropriate actions to control the robot in operating mode.

The main function of categorization networks is to categorize input S-patterns and A-patterns into S-categories and A-categories, respectively, for sensorimotor categorical mappings formation. Fuzzy ART network (Carpenter et al., 1991aa) is chosen to realize categorization networks in the BLOM architecture. The reason for choosing ART network to realize categorization networks is that they solve the *Stability–Plasticity Dilemma* which is faced by other categorizing and learning systems (Carpenter & Grossberg, 1987a). The learned categorization codes are stable to resist the erosion of irrelevant data while sensitive to novel data. These features allow ART networks to be a successful candidate for incremental categorization learning networks, which is an important property of on-line robot learning. Other neural network architectures designed specifically for classification or clustering, like LVQ (Kohonen, 1995), and statistical clustering techniques are either supervised in

[3] $a \tilde{\subset} S$ is defined as that a is a category or cluster in A.

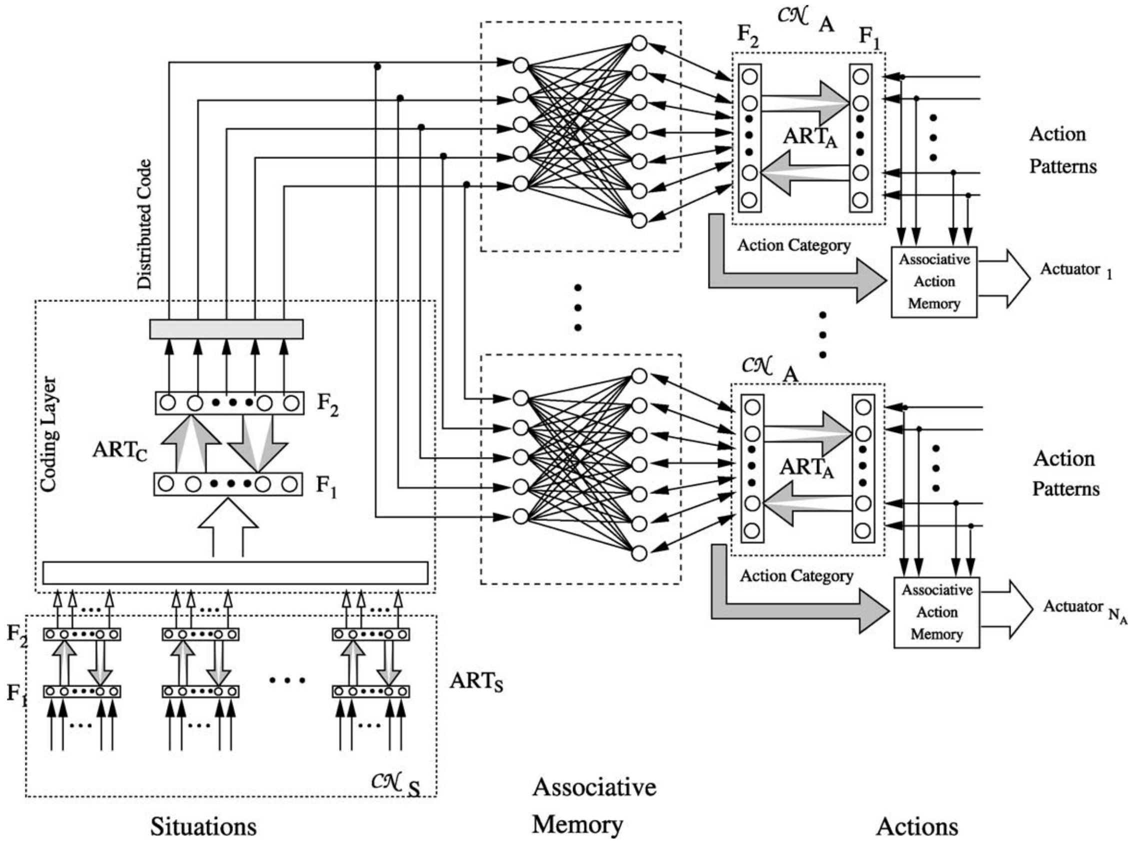


Fig. 6. The BLOM Architecture.

nature or not suitable for incremental learning. Moreover, the granularity of the clusters (categories) formed by ART networks can be controlled by vigilance parameter. The proposed adaptive categorization mechanism basically adjusts the vigilance parameters of the F_2 neurons to control the granularity of the categories formed automatically. As shown in the simulations in Section 3.5, the proposed game-theoretic adaptive categorization for ART networks improve the category number stability of ART networks and thus solve the problem of initial vigilance parameter prior to pattern categorization in traditional ART networks for robot behavior learning using the BLOM architecture. The selection of initial vigilance parameters for ART networks depends heavily on the distribution and other characteristics of the population of the patterns for categorization, which are unknown a priori. Extensive experiments have to be conducted to determine the suitable vigilance parameter for pattern categorization in trial-and-error basis before launching the ART networks for practical uses. With the introduction of adaptive categorization mechanism to ART networks in the BLOM architecture, the initial vigilance parameters can be arbitrarily set and allow the networks adapt to the patterns encountered in real-time. The adaptive categorization mechanism also effectively suppress the unnecessary granularity of categories formed in categorization in the cases with high vigilance parameters. The adaptive categorization mechanism place the balance well between the stability and plasticity of categorization networks. Therefore, the need of memory and computational power increase drastically during robot behavior learning if the adaptive categorization mechanism is not activated in the ART networks in the BLOM architecture. On the other hand, the adaptive categorization mechanism of ART networks can maintain slow increases of categories formed, and thus the need of memory and computational power during behavior learning without sacrificing the learning performance of the BLOM architecture. In additions, Fuzzy ART network can handle both binary and analog input patterns so that flexibility in data pattern encoding for behavior learning is enhanced.

4.2. Behavior learning experiments

Robot behavior learning experiments are conducted with a RWI B21 mobile robot using the BLOM architecture. Fig. 7 depicts a picture of the B21 mobile robot. The robot can perform holonomic motion so that motion planning for the robot is simplified. There are 24 ultrasonic range sensors, together with 56 infra-red sensors and 56 tactile switches, installed evenly on

Fig. 7. The RWI B21 mobile robot employed in the experiments.

the peripheral of the robot. The robot is also equipped with a stereo vision system. In the experiment, the mobile robot is employed to learn the *Wall-following* behavior from scratch. The Wall-following behavior guides a robot to move along a wall or boundary of objects in an environment while keeping a certain fixed distance from the wall or objects. Behavior learning is conducted in the *logical* perceptual space, which is extracted from the physical perceptual space of the robot using factor analysis (Fung & Liu, 2000). The logical perceptual space can be described by the measurement model extracted by factor analysis. The sensor data patterns, or S-patterns, involved in the experiments are constructed from 24 ultrasonic range data and its four time history values returned from the robot, which is the physical perceptual space S. The physical perceptual space is thus a 24D space and 11 latent factors can be extracted out of the physical perceptual space (Fung & Liu, 2000). Thus, there are 11 Fuzzy ART networks for S-patterns categorization with logical perceptual space training as each dimension of the logical perceptual space is associated with one Fuzzy ART network. Each Fuzzy ART network has 5D input patterns for categorization (current value plus its previous values at four consecutive time instances). On the other hand, the action patterns, or A-patterns, are constructed from the translational and rotational velocities of the robot and there are two Fuzzy ART networks, OLAM for A-patterns categorization and Action Memory in the BLOM architecture. Each ART network in the action side also has 5D A-patterns as input.

The experiments conducted can be divided into two phases, namely the Learning phase and Operating phase. In the Learning phase, human operator controls the B21 robot to exhibit the 'Wall-following' behavior that the robot is going to learn in training environments using the developed Tele-Assisted Teaching System (TATS), as shown in Fig. 8. TATS is a software for sensor-action pattern pairs acquisition for robot behavior learning with the proposed BLOM architecture. TATS acts as a *human–robot interface* that provides online sensory feedback visually to human operator for controlling robot motion. The online sensory feedback include local *sonar map* and images captured from the cameras equipped on the robot. This teaching system allows human 'teacher' to guide the motion of the mobile robot (A-patterns) by observing different sensor sources, including sonar range data and captured images (S-patterns). During 'teaching' the robot

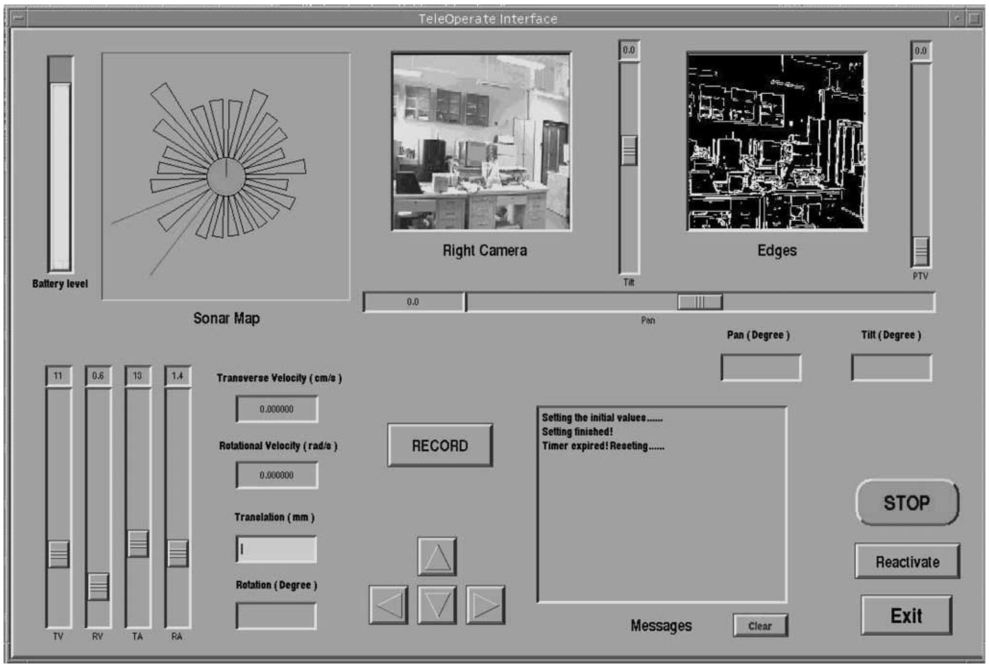


Fig. 8. Screenshot of the Tele-Assisted Teaching System (TATS).

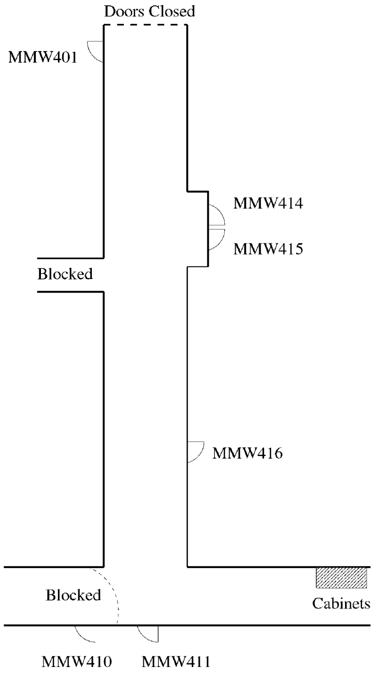


Fig. 9. A sketch of the corridor environment.

Fig. 10. Photos of the corridor environment.

with the TATS, S-patterns and A-patterns pairs are acquired and saved by the TATS. The robot is taught to follow the left wall and keep a fixed distance of about 20 cm from the wall or any object boundary in the corridors of the 4/F of Mong Man Wai Building, as shown in Figs. 9 and 10. Teaching lessons are repeated with different starting situations and environments until enough training patterns have been collected. There are 12,573 pairs of S-patterns and A-patterns collected in the training environments. Logical S-patterns are first generated from physical S-patterns and the measurement model obtained from Factor Analysis (Fung & Liu, 2000). Sensor (logical) and action training pattern pairs are then fed to the BLOM architecture to learn the 'Wall-following' behavior. Table 3 lists the number of categories (F_2 neurons) formed in the Fuzzy ART network of the BLOM architecture during robot behavior learning with and without applying the proposed vigilance parameter adaptation mechanism. The vigilance parameters of the Fuzzy ART networks are fixed at 0.87 when ρ-adaptation is not employed in robot behavior learning while the initial vigilance parameters of Fuzzy ART networks is set at 0.8 when ρ-adaptation is employed in robot behavior learning. As shown in Table 3, the number of categories formed in the Fuzzy ART networks with ρ-adaptation are less than one-fifth of the case when ρ-adaptation is not employed in pattern categorization for robot behavior learning and thus the effectiveness of the proposed game-theoretic adaptive categorization mechanism for ART networks is demonstrated.

After learning, the B21 robot is situated in novel environments and tested whether it can demonstrate the learned 'Wall-following' behavior in the Operating phase. The learned BLOM architecture is then set up in operating mode and it drives robot motion to operate in the novel environment to demonstrate the learned behavior. The input physical S-patterns are first transformed into corresponding logical S-patterns based on the measurement model for the BLOM architecture in operating mode. The robot is then tested to perform wall-following in the long and straight corridor outside MMW 410 and MMW 411. Fig. 11 depicts the trajectory of the robot guided by BLOM architecture learned with logical perceptual space and the map built when the robot performs wall-following. The robot follows the left wall in the corridor but the distance from the wall cannot be kept to a similar value to that of trained.

Table 3
Number of categories (two neurons) formed in Fuzzy ART (FART) networks after robot behavior learning in the logical perceptual dimension using the BLOM architecture

FART$_i$	1	2	3	4	5	6	7	8	9	10	11
w/ρ-adj.	487	515	508	542	458	602	572	496	531	576	553
w/o ρ-adj.	2783	2917	3069	2982	2906	3185	2684	2843	2939	2760	2883

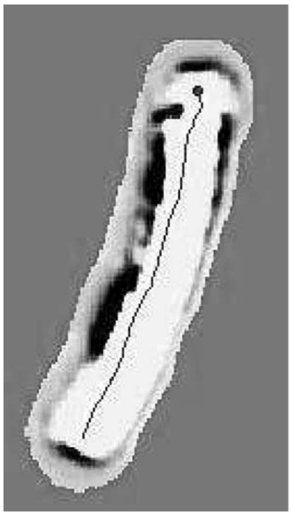


Fig. 11. Wall-following in Corridor-like environment.

The robot is then tested to perform wall-following in the 4/F lift lobby of Mong Man Wai Building, as shown in Figs. 12 and 13. Fig. 14 depicts the trajectory of the robot guided by BLOM architecture learned with logical perceptual space and the map built when the robot performs wall-following. The robot starts at the open free region in the lift lobby and they perform wall-following in different parts of the lift lobby.

5. Conclusions

This paper proposed a mathematical formulation of adaptive categorization of ART networks based on the game theory for robot behavior learning. We have derived

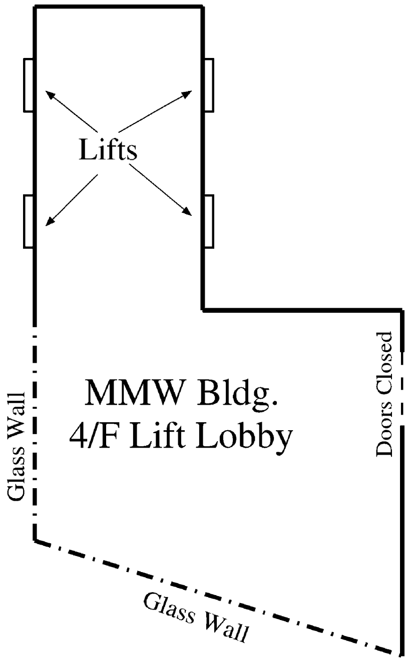


Fig. 12. A sketch of the lift lobby.

Fig. 13. Photos of the lift lobby.

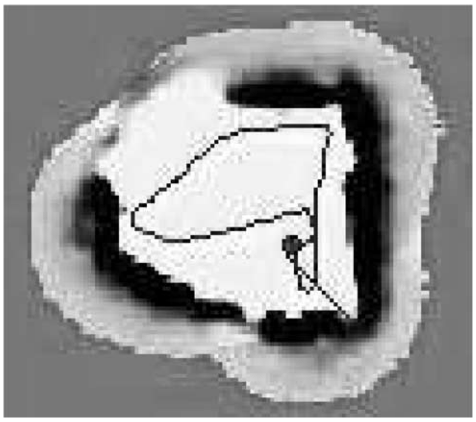


Fig. 14. Wall-following in Lift lobby.

the game-theoretic model Γ_{AC} for competitive processes of clustering of ART networks and an update rule for vigilance parameters using the concept of learning automata. Numbers of clusters generated by ART adaptive categorization are similar regardless of the initial vigilance parameters ρ assigned to the ART networks. The ρ-adaptation, thus, helps to solve the difficult problem of choosing suitable vigilance parameter prior to data categorization process and ease the design of BLOM architecture. Moreover, the coverage of clusters generated by ART networks with ρ-adaptation can reflect the shape of pattern distribution and thus prevent misclassification in classifiers constructed with ART networks. This phenomenon can sometimes be achieved by fixed high ρ in ART networks while ART networks with ρ-adaptation can achieve the same results even with low initial vigilance parameter. The proposed ART adaptive categorization mechanism can also avoid the problem of choosing suitable vigilance parameter a priori for pattern categorization. We also perform a repeated game analysis on the game Γ_{AC} by investigating the asymptotic behaviors of the update law for state probabilities and hence vigilance parameters. Several clustering experiments demonstrated that game-theoretic vigilance parameter adaptation can improve the clustering performance of ART networks in the aspect of category number stability. The stability of category number formed by ART categorization helps the architectural design of BLOM architecture. Robot behavior learning experiments are also conducted to demonstrate the effectiveness of the proposed adaptive categorization mechanism.

Acknowledgements

The authors would like thank the anonymous reviewers for their invaluable suggestions to the paper. This work is supported in part by the Hong Kong RGC through grants CUHK 415/97E, CUHK 4166/98E, CUHK 4173/00E and CUHK 404/01E.

Appendix

This section presents the proofs of the properties of the proposed adaptive categorization algorithm.

Lemma 1. *The proposed reinforcement scheme* (12) or the Markov process $\{\xi(t)\}_{t>0}$ is *strictly distance diminishing*.

Proof. Consider any two starting values of $\xi_u(t)$, namely p_1 and p_2, and their corresponding value of $\xi_u(t+1)$ as p'_1 and p'_2, respectively. From Eq. (12), we have

$$p'-p=\begin{cases}-\frac{1}{2}a(1+\theta)p+\frac{1}{2}b(1-\theta)(\frac{1}{2}-p) & \text{if } s(t)\neq\sigma_u\\ \frac{1}{2}a(1+\theta)(1-p)-\frac{1}{2}b(1-\theta)p & \text{if } s(t)=\sigma_u\end{cases}$$

with $p\in\{p_1,p_2\}$ and

$$\|p'_1-p'_2\|=\left(1-\frac{a+b}{2}-\frac{\theta(a-b)}{2}\right)\|p_1-p_2\|.$$

This implies that $\|p'_1-p'_2\|<\|p_1-p_2\|$ for $\theta\in\{-1,0,1\}$ and $0<a<1$, $0<b<1$ and hence $\{\xi(t)\}_{t>0}$ is strictly distance diminishing. □

Lemma 2. *The Markov process $\{\xi(t)\}_{t>0}$ has absorbing state with probability 1 if and only if $b=0$.*

Proof. (1) *Sufficiency*. If $b=0$, the reinforcement scheme becomes

$$\xi_u(t+1)=\begin{cases}\xi_u(t)-\frac{1}{2}a(1+\theta)\xi_u(t), & \text{if } s(t)\neq\sigma_u\\ \xi_u(t)+\frac{1}{2}a(1+\theta)(1-\xi_u(t) & \text{if } s(t)=\sigma_u\end{cases}$$

By setting

$$\xi_u(t+1)=\xi_u(t),$$

we have

$$\begin{cases}\xi_u(t)=0, & \text{if } s(t)\neq\sigma_u\\ \xi_u(t)=1 & \text{if } s(t)=\sigma_u\end{cases}$$

which is an absorbing state with probability 1 or state probability vector $\lim_{t\to\infty}\xi(t)$ is a unit vector.
(2) *Necessity*. If $\xi_u(t+1)=\xi_u(t)$, we have

$$\begin{cases}\xi_u(t)=\dfrac{(1/2)b(1-\theta)}{a(1+\theta)+b(1-\theta)}=0 & \text{if } s(t)\neq\sigma_u\\ \xi_u(t)=\dfrac{a(1+\theta)}{a(1+\theta)+b(1-\theta)}=1 & \text{if } s(t)=\sigma_u\end{cases}$$

because ξ_u can decrease only when σ_v because ξ_u can decrease only when $\sigma_v(\neq\sigma_u)$ is selected, which results in a favorable response $\theta=1$, and $\xi_u(t')$ is an unit vector if $\xi(t)$ is a unit vector $\forall t'>t$. This implies $b=0$.

Lemma 3. *The Markov process $\{\xi(t)\}_{t>0}$ is ergodic and has a limiting expected state probability $\xi^*\triangleq\lim_{t\to\infty}\mathbf{E}[\xi(t)]$ if and only if $b\neq0$.*

Proof. Let the environment characteristics be defined by the probabilities of $\theta(t)\in\{-1,0,1\}$ for states $s(t)\in\Sigma$,

$$\mathrm{Prob}(\theta(t)=-1|s(t)=\sigma_u)=\phi_u,\qquad 0<\phi_u<1$$
$$\mathrm{Prob}(\theta(t)=0|s(t)=\sigma_u)=\psi_u,\qquad 0<\psi_u<1$$
$$\mathrm{Prob}(\theta(t)=1|s(t)=\sigma_u)=1-\phi_u-\psi_u,\quad 0<\phi_u+\psi_u<1$$

(1) *Sufficiency*. In order to study the asymptotic behavior of $\{\xi(t)\}_{t>0}$, we first investigate the conditional expectation of $\xi(t+1)$ given $\xi(t)$. If $b\neq0$, we have

$$\begin{aligned}\Delta\xi_u(t)&\triangleq\mathbf{E}[\xi_u(t+1)-\xi_u(t)|\xi(t)]\\&=\xi_u(t)\sum_{\sigma_v\neq\sigma_u}\Big((1-\phi_u-\psi_u)g_v(\xi(t))-\phi_u h_v(\xi(t))\\&\quad+\frac{\psi_u}{2}[g_v(\xi(t))-h_v(\xi(t))]\Big)-\sum_{\sigma_v\neq\sigma_u}\xi_v(t)\Big((1-\phi_v-\psi_v)\\&\quad\times g_u(\xi(t))-\phi_v h_u(\xi(t))+\frac{\psi_v}{2}[g_u(\xi(t))-h_u(\xi(t))]\Big)\\&=-b\Big(\phi_u+\frac{\psi_u}{2}\Big)\xi_u(t)+\frac{b}{2}\Big(\phi_u+\frac{\psi_u}{2}\Big)\sum_{\sigma_v\neq\sigma_u}\xi_v(t)\end{aligned}\qquad\text{(A1)}$$

Taking expectation on both sides, we have

$$\mathbf{E}[\xi_u(t+1)] = \left[1 - b\left(\phi_u + \frac{\psi_u}{2}\right)\right]\mathbf{E}[\xi_u(t)] + \frac{b}{2}\left(\phi_u + \frac{\psi_u}{2}\right) \times \sum_{\sigma_v \neq \sigma_u} \mathbf{E}[\xi_v(t)] \quad \text{(A2)}$$

Combining equations with the form of Eq. (A2) for $\sigma_u \in \Sigma$, we have $\mathbf{E}[\xi(t+1)] = \mathbf{C}^{\mathrm{T}}\mathbf{E}[\xi(t)]$, where $\mathbf{C} \in \mathbb{R}^{3\times 3}$ and

$$\begin{cases} c_{uu} = 1 - b\left(\phi_u + \dfrac{\psi_u}{2}\right) \\ c_{uv} = \dfrac{b}{2}\left(\phi_u + \dfrac{\psi_u}{2}\right) \end{cases} \text{where } \sigma_v \neq \sigma_u$$

Since $c_{uv} \in (0,1)$, $\forall \sigma_u$, $\sigma_v \in \Sigma$ and $\sum_{\sigma_u \in \Sigma} c_{uv} = 1$, $\mathbf{C}$ is hence a stochastic matrix and all of its eigenvalues lie on or inside the unit circle. Therefore, the Markov process $\{\xi(t)\}_{t>0}$ is ergodic and the equilibrium point of the system is asymptotically stable. The ergodicity also implies that $\mathbf{E}[\xi(t)]$ has limiting value as $t \to \infty$ which is *independent* of the initial state probability $\xi(0)$. The limiting state probability ξ^* is the solution of $\xi^* = \mathbf{C}^{\mathrm{T}}\xi^*$ satisfying the $\sum_{\sigma_u \in \Sigma} \xi_u^* = 1$ constraint and is given as

$$\xi_u^* = \lim_{t\to\infty} \mathbf{E}[\xi_u(t)] = \frac{\dfrac{1}{\phi_u + \psi_u/2}}{\displaystyle\sum_{\sigma_v \in \Sigma} \dfrac{1}{\phi_v + \psi_v/2}}, \quad \text{where } \sigma_u \in \Sigma \quad \text{(A3)}$$

(2) *Necessity*. Suppose $b = 0$, the limiting expected state probability is an unit vector according to Lemma 2. Sinced $\{\xi(t)\}_{t>0}$ is ergodic, we have $\xi_u^* = 1$ and $\xi_v^* = 0$ for $\sigma_v \neq \sigma_u$. With Eq. (A3), this implies a contradiction, which states that $\sum_{\sigma_v \neq \sigma_u} \frac{1}{\phi_v + \frac{\psi_v}{2}} = 0$ or $\phi_v + \frac{\psi_v}{2} = -(\phi_w + \frac{\psi_w}{2}) < 0, \forall \sigma_v$, $\sigma_w \neq \sigma_u$, as ϕ_z^*, $\psi_z > 0$, $\forall \sigma_z \in \Sigma$. Therefore, $b \neq 0$ □

Lemma 4. *If $b \neq 0$, the learning automaton $\mathscr{L}_{AC}$ is expedient.*

Proof. A learning automaton is expedient if $\lim_{t\to\infty} \times \mathbf{E}[M(t)] > M_0$, where $M(t) \triangleq \mathbf{E}[\theta(t)|\xi(t)] = \sum_{\sigma_u \in \Sigma} x_u \xi_u(t)$ and $M_0 = (x_{\mathrm{R}} + x_{\mathrm{r}} + x_{\mathrm{f}})/3$ is the *norm of behavior* for pure-chance automaton (Narendra & Thathachar, 1989). From Lemmas 2 and 3 and since $x_u \in (-1,1)$ $\forall \sigma_u \in \Sigma$, we have

$$\lim_{t\to\infty} \mathbf{E}[M(t)] = \sum_{\sigma_u \in \Sigma} \left(\frac{\dfrac{x_u}{1-x_u}}{\displaystyle\sum_{\sigma_v \in \Sigma} \dfrac{1}{1-x_v}} \right) = \frac{(x_{\mathrm{R}} + x_{\mathrm{r}} + x_{\mathrm{f}}) - 2(x_{\mathrm{R}}x_{\mathrm{r}} + x_{\mathrm{R}}x_{\mathrm{f}} + x_{\mathrm{r}}x_{\mathrm{f}}) + 3x_{\mathrm{R}}x_{\mathrm{r}}x_{\mathrm{f}}}{3 - 2(x_{\mathrm{R}} + x_{\mathrm{r}} + x_{\mathrm{f}}) + (x_{\mathrm{R}}x_{\mathrm{r}} + x_{\mathrm{R}}x_{\mathrm{f}} + x_{\mathrm{r}}x_{\mathrm{f}})} > \frac{x_{\mathrm{R}} + x_{\mathrm{r}} + x_{\mathrm{f}}}{3} \quad \text{(A4)}$$

□

Lemma 5. *The learning automaton $\mathscr{L}_{AC}$ is absolutely expedient if and only if $b = 0$.*

Proof. A learning automaton is absolutely expedient if and only if all ratios $(g_u(\xi))/\xi_u$ are equal and all ratios $(h_u(\xi))/\xi_u$ are equal $\forall \sigma_u \in \Sigma$ (Narendra & Thathachar, 1989). From the proposed reinforcement scheme, the only solution that satisfies the above conditions is $b = 0$ such that $(g_u(\xi))/\xi_u = a$ and $(h_u(\xi))/\xi_u = 0$, $\forall \sigma_u \in \Sigma$. Hence the lemma is proved. □

References

Carpenter, G. A., & Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics and Image Processing*, *37*, 54–115.

Carpenter, G. A., & Grossberg, S. (1987b). ART2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, *260*, 4919–4930.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., & Rosen, D. B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multi-dimensional maps. *IEEE Transactions on Neural Networks*, *30*(5), 698–713.

Carpenter, G. A., Grossberg, S., & Rosen, B. (1991a). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, *4*, 759–771.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991b). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, *4*, 565–588.

Connell, J. H., & Mahadevan, S. (Eds.), (1993). *Robot learning*. Dordrecht: Kluwer Academic.

Franklin, J. A., Mitchell, T. M., & Thrun, S. (Eds.), (1996). *Recent advances in robot learning*. Dordrecht: Kluwer Academic.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: The MIT Press.

Fung, W. K., & Liu, H. (1998). *A behavior learning/operating module for mobile robots. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'98*, pp. 1879–1874.

Fung, W. K., & Liu, H. (2000). *Extracting Logical Perceptual Space for Robot Learning using Factor Analysis* (*Vol. 2*). *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'2000*, pp. 873–878.

Kohonen, T. (1989). *Self-organization and associative memory* (3rd ed). Berlin: Springer.

Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer.

Kosko, B. (1992). Fuzzy associative memory systems. In A. Kandel (Ed.), *Fuzzy Expert Systems* (pp. 135–162). Boca Raton: CRC Press.

Massaro, D. W. (1990). In O. Neumann, & W. Prinz (Eds.), *An Information-Processing Analysis of Perception and Action* (pp. 133–166). *Relationships Between Perception and Action Current Approaches*, Berlin: Spring.

Narendra, K. S., & Thathachar, M. A. L. (1989). *Learning automata: An introduction*. Englewood Cliffs, NJ: Prentice-Hall.

Perrberg, I. M. (1996). Categorical class formation by an African Grey Parrot (*Psittacus erithacus*). In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (*Vol. 117*) (pp. 71–90). *Advances in Psychology*, Amsterdam: Elsevier.

Vlajic, N., & Card, H. C. (1998). *Categorizing web pages using modified ART* (*Vol. 1*). *IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 313–316.

Walker, J. T. (1995). *The psychology of learning*. Englewood Cliffs, NJ: Prentice-Hall.

White, D. J. (1993). *Markov decision processes*. London: Wiley.

Williams, J. H., Jr (1996). *Fundamentals of applied dynamics*. London: Wiley.

Wai-keung Fung received the BE degree in electroni engineering, the MPhil degree in system engineering and engineering management, and the PhD degree in automation and computer aided engineering from the Chinese University of Hong Kong, Hong Kong, in 1994, 1996 and 2001, respectively. He was a postdoctoral research associate with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, in 2001–2002. He is now a postdoctoral fellow with the Department of Automation and Computer Engineering, The Chinese University of Hong Kong. His research interests include robot learning, intelligent robotics, services robotics, multiagents systems, computational intelligence and nanomanipulation.

Yun-hui Liu received his BE degree in applied dynamics from Beijing Institute of Technology, China, in 1985, the ME degree in mechanical engineering from Osaka University, Osaka, Japan, in 1989, and the PhD degree in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1992. He worked at the Electrotechnical Laboratory, MITI, Japan, as a Research Scientist from 1992 to 1995. Since February 1995, he has been with the Chinese University of Hong Kong and is currently a Professor in the Department of Automation and Computer Aided Engineering. Dr Liu is also a ChangJiang Professor of the National University of Defense Technology and the director of the Joint Center for Intelligent Sensing and Systems of the Chinese University of Hong Kong and the National University of Defense Technology. He has published over 100 papers in refereed journals and refereed conference proceedings. His research interests include adaptive force control, multi-fingered robotic hands, mobile robots, virtual reality, computer vision, and Internet-based robotic systems. Dr Liu received the 1994 and 1998 Best Paper Awards from the Robotics Society of Japan and the third place award for the outstanding paper of the IEEE First Electro/Information Technology Conference. He is an Associate Editor of IEEE Transactions on Robotics and Automation. He is the general chair of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).