

Retrograde adaptive resonance theory based on the role of nitric oxide in long-term potentiation

Peng Jia · Junsong Yin · Dewen Hu · Zongtan Zhou

Received: 24 July 2006 / Revised: 12 January 2007 / Accepted: 7 February 2007 / Published online: 3 April 2007
© Springer Science + Business Media, LLC 2007

Abstract Adaptive resonance theory (ART) demonstrates how the brain learns to recognize and categorize vast amounts of information by using top-down expectations and attentional focusing. ART 3, one member of the ART family, embeds the computational properties of the chemical synapse in its search process, but it converges slowly and is lack of stability when being applied in pattern recognition and analysis. To overcome these problems, Nitric Oxide (NO), which serves as a newly discovered retrograde messenger in Long-Term Potentiation (LTP), is introduced in retrograde adaptive resonance theory (ReART) model presented in this paper. In the presented model a novel search hypothesis is proposed to incorporate angle and amplitude information of an external input vector to decide whether the input matches the long-term memory (LTM) weights of an active node or not, and the embedded NO retrograde mechanism makes the search procedure a closed loop, which improves the stability and convergence speed of the transmitter releasing mechanism in a synapse. To make the model more adaptive and practical, a forgetting mechanism is built to improve the weights updating process. Experimental results indicate that the proposed ReART model achieves low error rate, fast convergence and self-organizing weights regulation.

Keywords Neural networks · Adaptive resonance theory · Retrograde ART (ReART) · Nitric oxide · Long-term potentiation · Forgetting mechanism

1 Introduction

Adaptive resonance theory (ART), introduced by Carpenter and Grossberg (2002), is a neural network that self-organizes recognition codes in response to sequences of input samples. It addresses the stability–plasticity dilemma successfully, which describes how a learning system can preserve its previously learned knowledge while keeping its ability to learn new patterns (Grossberg et al. 2004).

Research has led to the development of three classes of the ART systems: ART 1, ART 2 and ART 3. ART 1 self-organizes recognition categories in response to arbitrary sequences of binary input patterns (Carpenter and Grossberg 1987). ART 2 does the same for either binary or analog inputs. ART 3, based on ART 2, includes a model of the chemical synapse embedded in network hierarchies, which solves the memory search problem of ART systems (Carpenter and Grossberg 1990). Besides, variations of these models adapted to the demands of individual applications have been developed (Levine and Penz 1990; Kumar and Guez 1991; Williamson 1997; José and Wolfram 1999; Chang et al. 2000; Grossberg et al. 2004; Quah et al. 2005; Stephen et al. 2005; Feng and José 2006).

A basic ART system consists of two subsystems, an attentional subsystem and an orienting subsystem. The stabilization of learning and activation occurs in the attentional subsystem, which contains two layers, F_1 and F_2 shown in Fig. 1. Once the system receives an input vector, layer F_1 generates a bottom-up signal to layer F_2 . Then the system searches in its memory for the best-matched

Action Editor: Christiane Linster

P. Jia · J. Yin (✉) · D. Hu · Z. Zhou
College of Mechatronics and Automation,
National University of Defense Technology,
Changsha, Hunan 410073, China
e-mail: jsyin@nudt.edu.cn

D. Hu (✉)
e-mail: dwhu@nudt.edu.cn

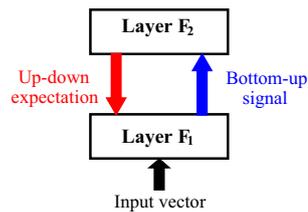


Fig. 1 Schematic diagram of attentional subsystem. Once the system receives an input vector, layer F_1 generates a bottom-up signal to layer F_2 . Then the system searches in its memory for the best-matched vector, and generates an up-down expectation signal to layer F_1 . This is called Resonance between the input vector and the relevant vector in memory. The matching criterion is given by a vigilance parameter ρ . If the bottom-up and top-down signals match each other, the relevant node's LTM weight will be updated. Otherwise, a reset signal generated by the orienting subsystem will lead to a new search-match cycle

category vector, activates the relevant node and generates an up-down expectation signal to layer F_1 . This is called Resonance between the input sample and the relevant node in memory. The matching criterion is given by a vigilance parameter ρ . Low vigilance allows broad generalization, coarse categories and abstract memories. High vigilance leads to narrow generalization, fine categories and detailed memories (Carpenter and Grossberg 2002). If the bottom-up and top-down signals match each other, the related node's long-term memory (LTM) weight will be updated. Otherwise, a reset signal generated by the orienting subsystem will lead to a new search-match cycle. That is to say that the orienting subsystem controls the attentional subsystem when a mismatch occurs in the attentional subsystem.

ART 3, one member in the ART family, can be either fast or slow learning and distributed of compressed code representations. Although some main properties of neurotransmitters are embedded in the architecture model, there are still some limitations in classification accuracy, convergence speed, and weights updating mechanism of ART 3, as well as ART 3 is sensitive to the selection of vigilance parameter; therefore, further research is needed and more properties of neurotransmitters can be introduced to ART 3 to overcome the shortcomings.

It is well known that Nitric Oxide (NO), as a neurotransmitter, plays a very important role in synaptic actions. The traditional neurotransmitters, such as Ca^{2+} and Na^+ , release from point to point. But NO, as a very small and nonpolar molecule, will transmit four-dimensionally in both space and time. It has strong neuro-modulating ability, which provides new inspiration about some of the traditional tenets in the domain of brain science, artificial neural network and so on. Experiments have shown that repetitive activation of excitatory synapses in the hippocampus, a brain region known well to be essential for learning and memory, caused an increase in synaptic strength that could last for hours or even days (Gladys and Kelly 1999). This

long-lasting synaptic enhancement, known as Long-Term Potentiation (LTP), is widely believed to be induced by learning (Whitlock et al. 2006), and provides an important key to understand the cellular and molecular mechanisms by which memories are formed and stored. NO plays the role of retrograde messenger in LTP (Christelle and Garthwaite 2003). It diffuses out of postsynaptic sites into presynaptic sites, leading to the increase of the neurotransmitter release.

Naturally embedding NO retrograde mechanism in ART3, this paper presents a novel adaptive resonance theory model named Retrograde ART (ReART in short). Based on NO retrograde mechanism in LTP, the novel model, ReART, presents an efficient search hypothesis, improves the presynaptic transmitter accumulation and release, thereby it achieves lower error rate, faster convergence in data classification than the original ART 3. Furthermore, because of the closed loop search procedure the embedded retrograde mechanism also improves the stability of the model. Another feature of ReART is its unique weights updating mechanism. By revisiting the characteristic of human brain memory (Solan and Ruppini 2001; Ochsner et al. 2002), we introduce a forgetting mechanism, which helps reducing memory consumption. Simulations demonstrate that the proposed model has self-organized weights regulation ability.

The rest of this paper is organized as follows. Section 2 first provides some materials about ART 3 chemical synapse model and NO retrograde mechanism in LTP, and then clarifies the proposed ReART network based on NO retrograde physiological mechanism. Section 3 shows the simulation results of ReART comparing with the original ART 3. The further analyses and discussions are presented in Section 4.

2 Methodologies

2.1 ART 3 chemical synapse

The electrical, ionic, and chemical events in a dendritic spine synapse are detailed by Carpenter and Grossberg (1990). The axon terminal produces neurotransmitter continuously. Some of them will be stored in the axon terminal. Some will be released to postsynaptic receptors, leading to the change of ionic permeability of postsynaptic membrane. Others may move to extra cellular space, or be destroyed by an enzyme, or absorbed by the presynaptic terminal.

The notation to describe these transmitter properties in ART 3 is summarized in Fig. 2 for a synapse between the i th presynaptic node and the j th postsynaptic node. The presynaptic signal, or action potential S_i arrives at a

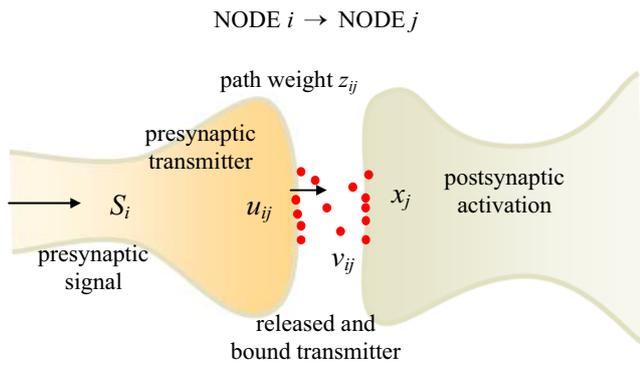


Fig. 2 Schematic drawing for ART 3 chemical synapse. When the presynaptic signal S_i arrives at a synapse whose long-term memory is denoted by z_{ij} , transmitter u_{ij} available for release is typically released. The variable v_{ij} denotes the amount of transmitter released into the extra cellular space. Finally, x_j denotes the activity of the postsynaptic cell

synapse whose adaptive weight, or long-term memory, is denoted by z_{ij} . The variable z_{ij} is identified with the maximum amount of available transmitter. When the transmitter at this synapse is fully accumulated, the amount of transmitter u_{ij} available for release is equal to z_{ij} . When a signal S_i arrives, transmitter is typically released. The variable v_{ij} denotes the amount of transmitter released into the extra cellular space, a fraction of which is assumed to be bound at the postsynaptic cell surface and the remainder rendered ineffective in the extra cellular space. Finally, x_j denotes the activity, or membrane potential, of the postsynaptic cell.

2.2 ART 3 dynamical equations

Equations (1–3) govern the dynamics of the variables z_{ij} , u_{ij} , v_{ij} at the ij th pathway and x_j for the j th node of an ART 3 system, where J_i denotes the excitatory inputs, and y_i the inhibitory inputs (Carpenter and Grossberg 1990).

$$\frac{du_{ij}}{dt} = (z_{ij} - u_{ij}) - u_{ij} [\text{release rate } \xi] \tag{1}$$

$$\frac{dv_{ij}}{dt} = -v_{ij} + u_{ij}\xi - v_{ij} [\text{reset signal } \zeta] \tag{2}$$

$$\begin{aligned} \varepsilon \frac{dx_i}{dt} &= -x_i + (A - x_i)J_i - (B + x_i)y_i \\ &= -x_i + (A - x_i) \times \left\{ \sum_j v_{ij} + \sum_l v_{il} \right\} - (B + x_i)\zeta \end{aligned} \tag{3}$$

Equation (1) says that presynaptic transmitter is produced or mobilized until the amount u_{ij} of transmitter available for release reaches the maximum level z_{ij} . The

adaptive weight z_{ij} itself changes on a slower time scale of learning, and remains essentially constant on the time scale of a single reset event. The ART search hypothesis I indicates that the available presynaptic transmitter u_{ij} is released at a rate that is proportional to $S_i f(x_j)$, where

$$f(x_j) = \begin{cases} 0, & x_j \leq -\varepsilon_1 \\ \frac{\varepsilon_2}{\varepsilon_1} \cdot x_j + \varepsilon_2, & x_j > -\varepsilon_1 \end{cases} \quad \varepsilon_1, \varepsilon_2 > 0.$$

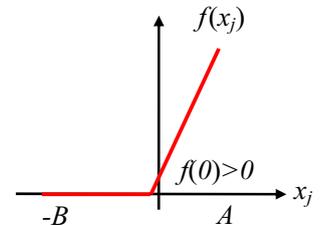
A fraction of the presynaptic transmitter becomes the postsynaptic bound transmitter after being released. For simplicity, we ignore the fraction of released transmitter that is inactivated in the extra cellular space. To implement the reset process, the ART search hypothesis II is postulated: the nonspecific reset signal quickly inactivates postsynaptic membrane channels at which the transmitter is bound. Equation (2) indicates that the bound transmitter is inactivated by the reset signal.

Equation (3), for the postsynaptic activity x_j , is a shunting membrane equation such that excitatory inputs drive x_j up toward a maximum depolarized level equal to A ; inhibitory inputs drive x_j down toward a minimum hyperpolarized level equal to $-B$; and activity passively decays to a resting level equal to 0 in the absence of inputs, as shown in Fig. 3. The net effect of bound transmitter at all synapses converging on the j th node is assumed to be excitatory, via the term: $\sum_l v_{il}$. Internal feedback from within the target field is excitatory, while the nonspecific reset signal is inhibitory. Parameter ε in Eq. (3) is small, corresponding to the assumption that activation dynamics are fast relative to the transmitter accumulation rate.

The ART 3 system can be simplified for purposes of simulation. Suppose that $\varepsilon \ll 1$ in Eq. (3); the reset signals in Eqs. (2) and (3) are either 0 or much larger than 1; and net intra-field feedback is excitatory. Then Eqs. (1), (4) and (5) approximate the main properties of ART 3 system dynamics.

$$\begin{cases} \frac{dv_{ij}}{dt} = -v_{ij} + u_{ij}\xi, & \zeta = 0 \\ v_{ij} = 0, & \zeta \gg 1 \end{cases} \tag{4}$$

Fig. 3 The nonlinear function $f(x_j)$. $f(0) > 0$ guarantees that transmitter can be released when the signal S arrives at the synapse



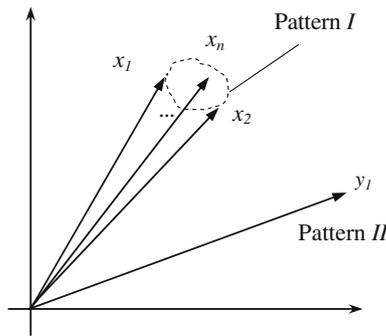


Fig. 4 The input sequences x_1, x_2, \dots, x_n are all belong to pattern I, and y_1 belongs to pattern II

$$\begin{cases} x_i = A \left(\sum_j v_{ij} + \sum_l v_{il} \right) / \left(1 + \sum_j v_{ij} + \sum_l v_{il} \right), & \zeta = 0 \\ x_i = 0, & \zeta \gg 1 \end{cases} \quad (5)$$

2.3 Analyses of ART 3

1) Miss of the perfect matched node

During the search process of ART 3, there may exist two or more nodes whose LTM weights Z_i match S well:

$$\cos(S, Z_i) \geq \rho \quad i = 1, 2, \dots, m, \quad (6)$$

where m denotes the number of nodes encoded in memory. But among Z_1, Z_2, \dots, Z_m , there is only one perfect match Z_I to S :

$$\cos(S, Z_I) \geq \cos(S, Z_i) \geq \rho \quad i \neq I. \quad (7)$$

In original ART 3, the I th node may not be first activated because the length $\|Z_I\|$ is small, then:

$$S \cdot Z_i \geq S \cdot Z_I. \quad (8)$$

So the perfectly matched node for the input pattern couldn't win first. The mismatched but with large weight value nodes may be undesirably preferred; probably the perfectly matched node wins only if other nodes' weights iteratively get even smaller values by the reset signal. The search-reset circle needs to be improved.

2) Repeated search and reset

In ART 3 search process, an input pattern may later activate a different category node I whose weights are large but whose vector of LTM weights forms a poorer match than the original, smaller weights. Then the reset signal will selectively inactivate transmitter at postsynaptic sites adjacent to the I th node. Following such a reset wave, the new signal pattern $S \cdot U_i$ will be biased against the I th node relative to the original pattern. However, it could happen that the time interval prior to

the reset signal is so brief that only a small fraction of available transmitter is released. Then $S \cdot U_i$ could still be large relative to a correct $S \cdot U_j$. If this were to happen, the I th node would simply be selected again, then reset again, leading to an accumulating bias against that choice in the next time interval. This process could continue until enough transmitters V_i is inactivated to allow another node, with smaller weights Z_i but a better pattern match, to win the competition. Sometimes the reset event may iterate so many times that it converges slowly.

3) The potential phenomenon of pattern excursion

The pattern matching process in ART system is to compare an external input vector with the internal LTM weights of an active node. The matching criterion is determined by the vigilance parameter ρ . It leads either to a resonant state, which persists long enough to permit learning, or to a parallel memory search. If the search ends at an established code, the LTM weights may either remain the same or incorporate new information from matched portions of the current input. If the search ends at a new node, new LTM weights

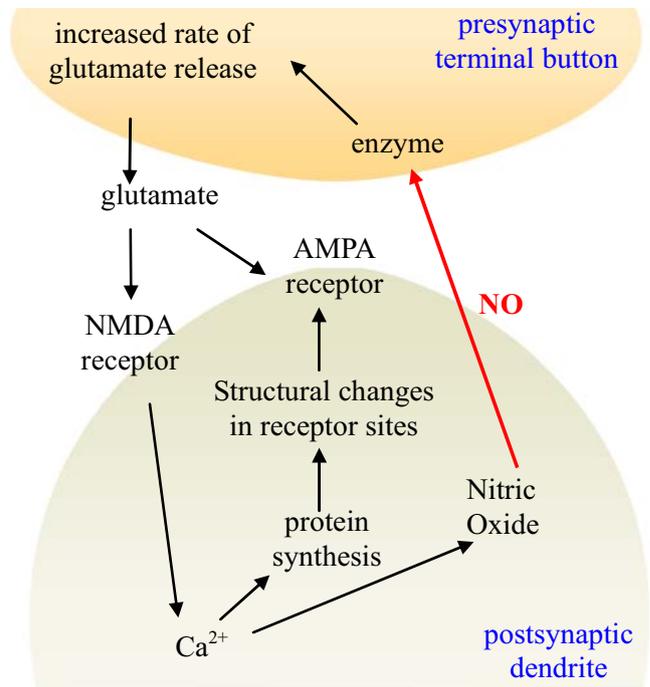


Fig. 5 Summary mechanism for LTP. Ca^{2+} channels open when glutamate activates NMDA receptors. If the membrane of the postsynaptic neuron is partially depolarized, the Ca^{2+} channels are neurotransmitter and voltage dependent, and more AMPA receptors are present. Then in order for the structure of the cell to change, new proteins must be synthesized. It is quite possible that Ca^{2+} channels act to initiate protein synthesis and to stimulate the diffusion of nitric oxide. As a soluble gas, NO could diffuse out of the postsynaptic membrane, and interact with an enzyme in the presynaptic terminal button, which is involved in the production of glutamate, resulting in an increase in the rate of glutamate release

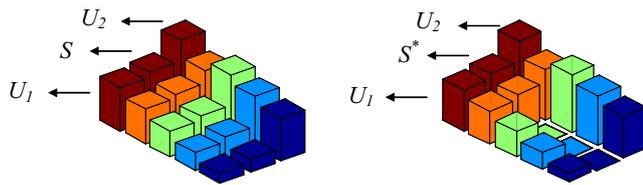


Fig. 6 Schematic drawing for U_1, S, U_2 (the left one) and U_1, S^*, U_2 (the right one)

learn the current input. Match-based learning allows memories to change only when input from the external world is close enough to internal expectations, or when something completely new occurs.

Assume the input sequences are x_1, x_2, \dots, x_n which all belong to pattern I . As illustrated in Fig. 4, the angles between them are:

$$\frac{x_1 \cdot x_2}{|x_1| \cdot |x_2|} < \rho \quad \frac{x_i \cdot x_j}{|x_i| \cdot |x_j|} > \rho, \quad i, j = 1, 3, 4, \dots, n. \quad (9)$$

When x_1, x_2, \dots, x_n are presented, the search and learning process is: first, node 1 will be encoded to learn x_1 :

$$w_1 = x_1. \quad (10)$$

Then switching to x_2 , which isn't close enough to x_1 , node 2 is encoded to learn x_2 ; third, input the sequences: x_3, \dots, x_n , they are close enough to x_1 , so node 1 is selected and modified. At last, there are two nodes to store pattern I . When the sequences are input repeatedly, there could be more and more nodes encoded to store the same pattern. Undoubtedly, it squanders the memory space.

In addition, when inputting y_1 , one node is encoded to learn pattern II . But y_1 is the only vector that belongs to pattern II , so pattern II is not representative, or even a noise vector. Still, the network holds one node to store it. This also squanders the memory space, and even causes wrong clustering. We call this the potential phenomenon of pattern excursion.

The first two shortcomings are partly because the transmitter release mechanism between the embedded chemical synapse is not quite complete. Fortunately, NO

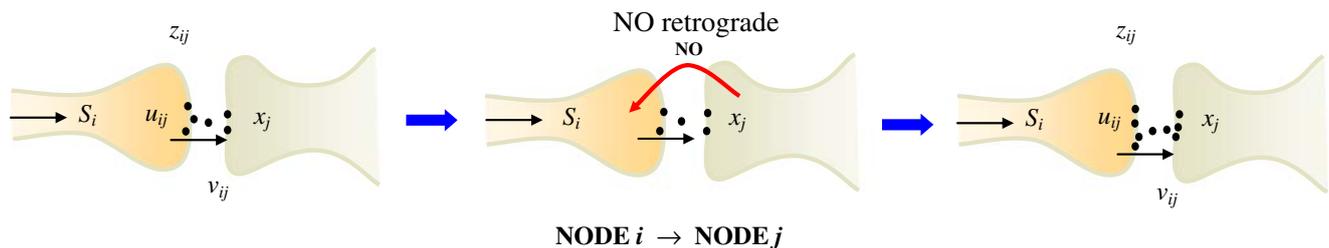


Fig. 7 Increase of transmitter release based on NO retrograde. When the presynaptic signal S_i arrives at a synapse, transmitter is accumulated and released (left one). The transmitter bound at the postsynaptic cell stimulates

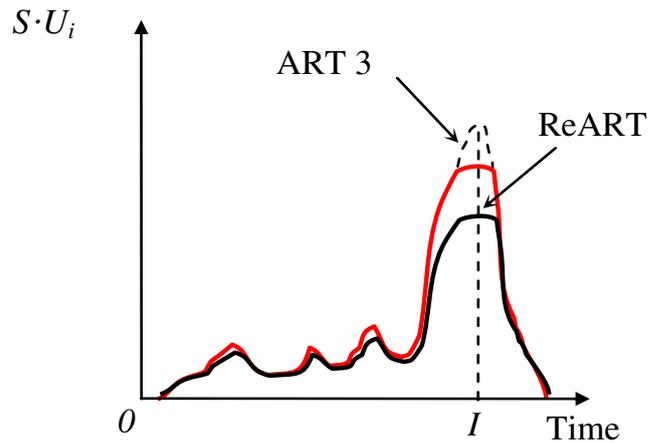


Fig. 8 The decrease of the functional value U_i during the reset events. The reset wave can decrease the mismatch weights' functional value

has special neuro-modulating ability, especially serves as retrograde messenger in LTP, which provides new inspiration for transmitter release mechanism between synapse.

2.4 Nitric oxide in long-term potentiation

A role for NO as an intercellular signaling molecule in the nervous system was first suggested by Garthwaite et al. (1988). NO is a very small and nonpolar molecule, and it will spread in three dimensions away from a site of synthesis regardless of intervening cellular or membrane structures. Another feature of NO signaling is suggested by the fact that the neuronal isoform of nitric oxide synthase (nNOS) is a soluble enzyme and likely to be distributed throughout a neuron's cytoplasm. The whole surface of the neuron is therefore a potential release site for NO. These properties allow NO to act without the need for presynaptic specializations, and its action is not necessarily confined to the immediate postsynaptic neuron. In conclusion, NO has opened new dimensions in our thinking about how information is transmitted by neurons (Bredt and Snyder 1992; Malenka and Nicoll 1999; Philippides et al. 2000; Rosenberg et al. 2000; Smith and Philippides 2000; Burette et al. 2002; Yang 2003; Ledo et al. 2005).

Table 1 Parameters used for the simulations

$$u_{ij} = \begin{cases} 0, & x_j \leq -0.01 \\ S_i \cdot (3 \cdot x_j + 0.1), & x_j > -0.01 \end{cases}$$

$$e = 0.001$$

$$d = 0.85$$

The available presynaptic transmitter u_{ij} is concerned with the presynaptic signal S_i and the activity of the postsynaptic cell, x_j . The parameter e is small. The LTM weights are updated at the rate of d .

In neuroscience, Long-Term Potentiation is the potentiation of the connection between two nerve cells, which lasts for an extended period of time (minutes to hours *in vitro* and hours to days or even months *in vivo*). LTP can be induced experimentally by applying a sequence of short, high frequency stimulations (HFS) to nerve cell synapses. The phenomenon was discovered in the hippocampus over 30 years ago. Recent research shows that one-trial inhibitory avoidance learning in rats produced the same changes in hippocampal glutamate receptors as induction of LTP with HFS and concludes that learning actually induces LTP in hippocampal area CA1 (Whitlock et al. 2006).

The summary mechanism for LTP is illustrated in Fig. 5. Glutamate (also called glutamic acid) is the most common excitatory neurotransmitter in the central nervous system, and NMDA is a type of glutamate receptor. Researches indicate that Ca^{2+} channels, which are normally blocked by Mg^{2+} , open when glutamate activates NMDA receptors. If the membrane of the postsynaptic neuron is partially depolarized, the Ca^{2+} channels are neurotransmitter and voltage dependent, and more AMPA, another type of glutamate receptor, are present. Then in order for the structure of the cell to change, new proteins must be synthesized. It is quite possible that Ca^{2+} channels act to initiate protein synthesis and to stimulate the diffusion of nitric oxide. As a soluble gas, NO could diffuse out of the postsynaptic membrane, and interact with an enzyme in the presynaptic terminal button, which is involved in the production of glutamate, resulting in an increase in the rate

of glutamate release. Research also indicates that postsynaptic Ca^{2+} influx through glutamate receptors and subsequent postsynaptic vesicle fusion trigger a robust induction of presynaptic miniature release after high-frequency stimulation at *Drosophila* neuromuscular junctions (Yoshihara et al. 2005).

In one word, during LTP, NO diffuses out of postsynaptic sites into presynaptic sites, leading to the increase of neurotransmitter release and improving the neurotransmitting mechanism, which gives inspiration to the original ART 3 neural networks.

2.5 The retrograde ART search model

In this section, we propose a novel retrograde ART search model. ReART search model consists of a new search mechanism based on NO retrograde in LTP and the node pruning mechanism.

2.5.1 A novel search hypothesis

ART search hypothesis III: if the intracellular transmitter U_i isn't the perfect match for the signal vector S , the amount of transmitter that will be released from the i th node to its postsynaptic node is decreased.

We know that the angle between S and U_i can measure the pattern match degree:

$$\cos(S, U_i) = \frac{S \cdot U_i}{\|S\| \cdot \|U_i\|}$$

Define a vector $\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2, \dots, \bar{\sigma}_n)$ which can represent the angle information between $S=(s_1, s_2, \dots, s_n)$ and $U_i=(u_{i1}, u_{i2}, \dots, u_{in})$ to a great extent:

$$\bar{\sigma}_k = \frac{|s_k - u_{ik}|}{s_k + e} \quad k = 1, 2, \dots, n, \tag{11}$$

where e is small and positive, which avoids dividing by zero. $\bar{\sigma}_k$ measures the similarity between s_k and u_{ik} .

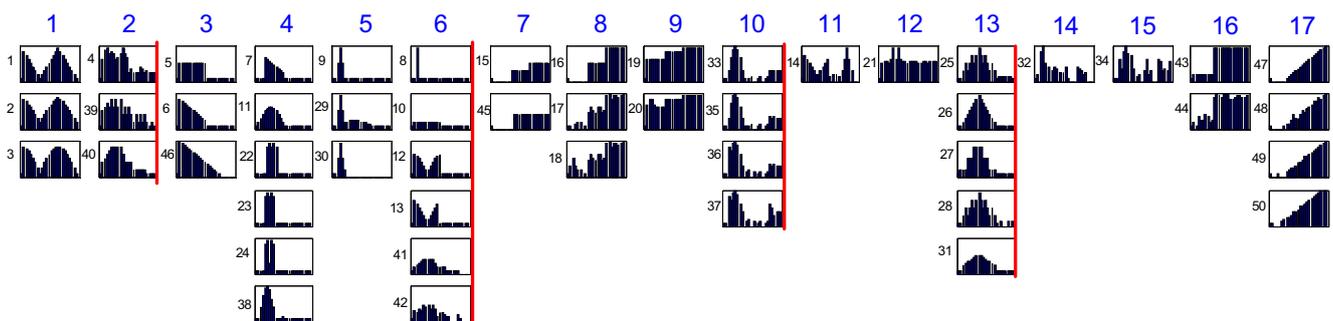


Fig. 9 The results of ART 3, $\rho=0.95$. ART 3 sorts the input patterns into 17 categories

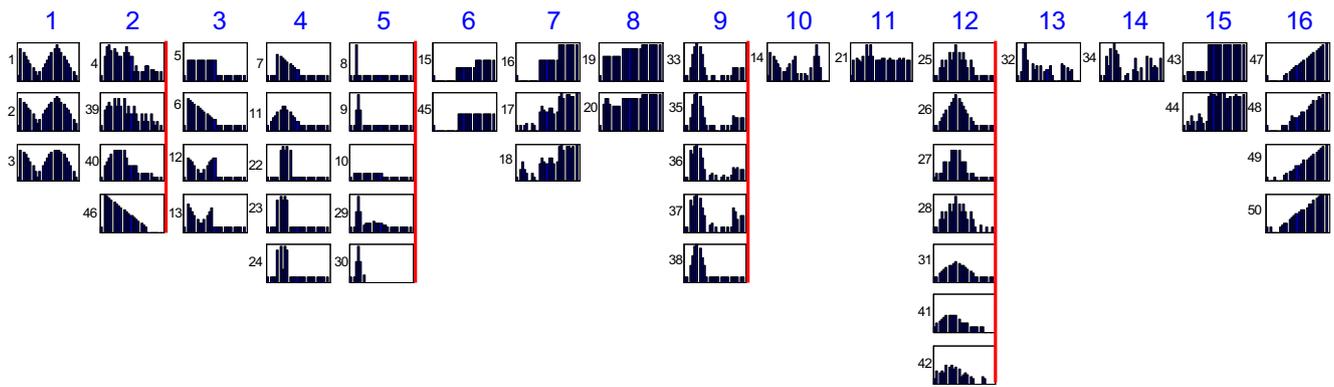


Fig. 10 The results of ReART, $\rho=0.95$, $\rho^*=0.5$. ReART sorts the 50 patterns into 16 categories: pattern 46 appears in category 2 instead of

category 3; category 12 won in response to inputs 41 and 42; pattern 38 was placed in category 9; the third and fifth categories obtained better groups

Compare $\bar{\sigma}_k$ with vigilance parameter ρ^* , where $\rho^* \in (0,1)$, which is usually set equal to 0.5. We get a novel input vector S^* :

$$s_k^* = \begin{cases} s_k, & \sigma_k < \rho^* \\ 0, & \sigma_k \geq \rho^* \end{cases} \quad (12)$$

That means, if s_k is similar to u_{ik} for $k = 1, 2, \dots, n$, we consider U_i matches S well; otherwise, U_i isn't the perfect match to S , and there will be less transmitter released to i th postsynaptic node according to search hypothesis III.

The calculating process is just like the “and” operation in Boolean calculation. We give a simple example here. Suppose the input vector $S = [1.5, 2.5, 3.5, 4.5, 5.5]^T$, and there are two nodes stored in the system whose transmitter available for release are: $U_1 = [1.1, 2, 3, 4, 5]^T$, $U_2 = [5, 6, 7, 6, 8]^T$, as shown in Fig. 6. Let $\rho^* = 0.5$, $e = 0.001$, then we get the following inequations:

$$\frac{|s_k - u_{1k}|}{s_k + 0.001} \rho^* = 0.5, \quad k = 1, 2, \dots, 5$$

$$\frac{|s_k - u_{2k}|}{s_k + 0.001} > \rho^* = 0.5, \quad k = 1, 2, 3$$

$$\frac{|s_k - u_{2k}|}{s_k + 0.001} < \rho^* = 0.5, \quad k = 4, 5$$

So we have the new input $S^* = [0, 0, 0, 4.5, 5.5]$ as seen in Fig. 6, which represents the match degree between input signal and the LTM weights of an active node.

S^* , instead of S , serves as the presynaptic signal, containing more information of patterns match degree.

Simulations demonstrate the efficacy and accuracy of the proposed model.

2.5.2 Increase of transmitter releasing based on NO retrograde

As we mentioned earlier, NO diffuses out of postsynaptic sites into presynaptic sites in LTP, increasing neurotransmitter release. Based on NO retrograde mechanism, we rebuild the release equation in ART 3:

$$\frac{du_{ij}}{dt} = (z_{ij} - u_{ij}) - u_{ij}\xi + u_{ij}x_j \quad (13)$$

The increment $u_{ij}x_j$, as a result of NO retrograde, is proportional to the postsynaptic activity x_j . Equation (13) indicates that thanks to NO retrograde, there will be more transmitters released to the postsynaptic sites in each time interval shown in Fig. 7, so the reset wave can decrease the mismatch weights' functional value u_{ij} much more than original ART 3, as illustrated in Fig. 8. Therefore, the improved model converges faster in simulations on conventional computers.

2.6 The forgetting mechanism

2.6.1 The forgetting mechanism

ART 3 network model is unsupervised. Along with the learning process of the input vectors, the weights in memory will be modified repeatedly, and there will be more and more nodes encoded in memory. However, the

Table 2 Comparison of recognition rate

Algorithm	SOM	K-means	ART 3	ReART
Recognition rate (%)	76.0	80.0	80.0	88.0

Fig. 11 Input samples

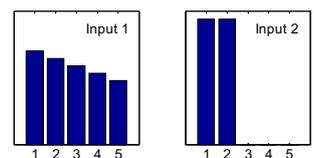
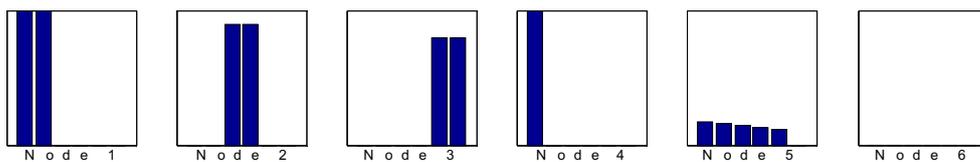


Fig. 12 LTM weights



input vectors are random, or even wrong, to avoid the potential phenomenon of pattern excursion we construct the forgetting mechanism, which is one of the great functions of human brain. The model is to define a cumulating matrix N for each node in memory to record the number of a node being learned, whose initial value is equal to 0. When the i th node is learned once, $N(i)$ adds 1:

$$N(i) = N(i) + 1. \tag{14}$$

Checking all the nodes in memory every stretch T (called forgetting interval), in case:

- 1) There exist two nodes whose weights satisfy:

$$\frac{w_i \cdot w_j}{|w_i| \cdot |w_j|} \geq \rho, \quad i \neq j \quad i, j = 1, \dots, n, \tag{15}$$

incorporate the i th and j th node to encode a new node with its LTM weights w :

$$w = \frac{w_i N(i)}{[N(i) + N(j)]} + \frac{w_j N(j)}{[N(i) + N(j)]}. \tag{16}$$

In other words, the network may generate two or more different nodes to learn the same category, and then we need to prune the excrescent nodes.

- 2) The number of input vectors is very large, we check matrix N in each forgetting interval T , when

$$N(I) \leq n^*, \tag{17}$$

we consider that the I th node is an erroneous node, abandon it, where vigilance n^* is a positive integer number, whose value is according to the number of input patterns. It means

that if a node is learned less than n^* times, we think it to be an erroneous node, or noise vector, which should be pruned.

The forgetting mechanism, which incorporates the matched nodes and prunes the erroneous nodes, regulates the LTM weights dynamically, and efficiently makes use of the memory space. The general node forgetting adjustment process is:

ReART Model

Input: input sample X , vigilance ρ , match degree vigilance ρ^* , learning times vigilance n^*

Output: LTM weights W

- Step 1. Data pretreatment;
- Step 2. Compete for a winner node I ;
- Step 3. Check for reset;
- Step 4. If the winner node matches the current input vector by vigilance ρ , then update LTM weights, or else reset U_I values and go to step 2;
- Step 5. If the current time gets to the forgetting interval T , then do the node forgetting adjustment, or else input the next pattern and go to step 1;
- Step 6. End.

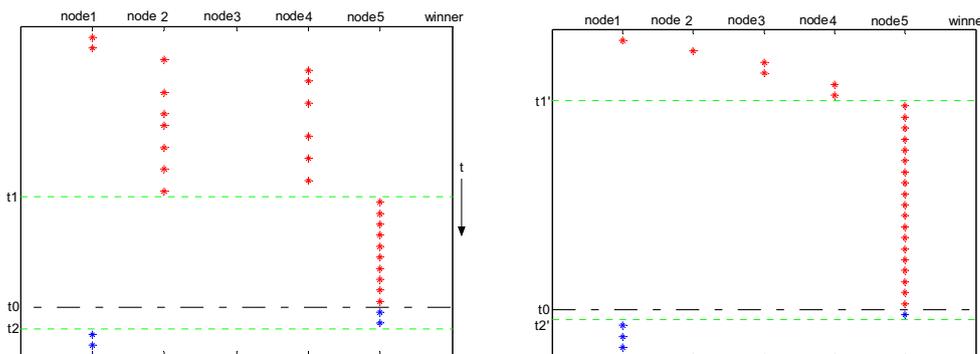
3 Results

In this section, the performance evaluation of ReART is given. Table 1 shows the parameters used for the simulations.

3.1 Comparison of accuracy on analog input patterns

The simulations summarized in Figs. 9 and 10 illustrate how ART 3 and ReART group 50 analog input patterns. There are 25 discriminating variables in each input vector.

Fig. 13 The winning nodes in each search-reset cycle, where $\rho=0.98$. A series of mismatch resets lead to the matched category. ART 3 activate the matched category at $t=t1$ (left one). ReART comparatively causes less resets, and find the matched category at $t=t1'$, Obviously, $t1 > t1'$ (right one). At $t=t0$, input 1 switches to input 2. Also, the searching time $(t2-t0) > (t2'-t0)$



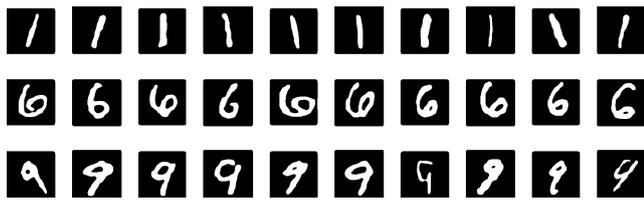


Fig. 14 Input digits images of “1,” “6” and “9”

The samples were repeatedly presented in the order 1,2,...,50 until the category structure stabilized. ART 3 sorts the 50 patterns into 17 recognition categories, while ReART sorts them into 16 categories. We can see some delighted improvement in Fig. 10: pattern 46 appears in category 2 instead of category 3; category 12 won in response to inputs 41 and 42; pattern 38 was placed in category 9; the third and fifth categories obtained better groups. In other words, more input patterns find its perfect match category.

The improved classification of ReART is due, in part, to the introduction of the magnitude information. During the search process, when a presynaptic signal S_i arrived at the synapse, it firstly compares with u_{ij} according to ART search hypothesis III. If the amplitude information of S_i matches u_{ij} by threshold ρ^* , nothing will be done with S_i ; or else S_i should be decreased properly. So there is less transmitter bound at the postsynaptic cell surface whose weights in the bottom-up pathway don't match S_i well, and the perfect matched node could win first in the competition.

We also sort the 50 input patterns into 16 categories by using self-organizing maps (SOM) (Kohonen 1997) and K -means algorithm (MacQueen 1967) using MATLAB software. Here is the comparison of the recognition rate in Table 2, which indicates that ReART carries out the best classification.

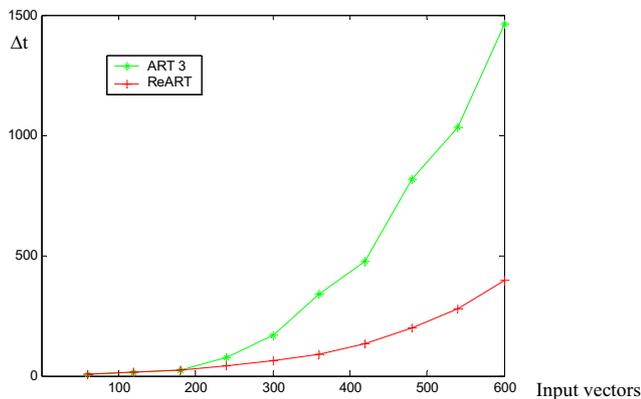


Fig. 15 Time consumption curve. Thanks to NO retrograde, ReART runs faster than ART 3

Table 3 Distribution of Iris data

Attribute	Range in mm
Sepal length	[43,79]
Sepal width	[22,44]
Petal length	[10,69]
Petal width	[1,25]

3.2 Comparison of time consumption

3.2.1 Six-dimensional input vectors

The input vectors and the bottom-up LTM weights shown in Figs. 11 and 12 are taken from Carpenter and Grossberg (1990). The input set consists of two patterns, which contains six discriminating variables. There are already five nodes encoded in memory.

Ignoring magnitude information, we use ART 3 and ReART to implement the classification. The winning nodes in each search-reset cycle are shown in Fig. 13. The introduction of ART search hypothesis III makes the winning nodes in a different time sequence, and new nodes (node 3) have chance of winning in the competition.

3.2.2 Handwritten digits

This example illustrates how ART 3 and ReART cluster a set of handwritten digits. The dataset “MNIST handwritten digits” is available at <http://www.cs.toronto.edu/~roweis/data.html>. It contains grayscale 28×28 digits of “0” through “9”; about 6,000 training examples of each class; 1,000 test

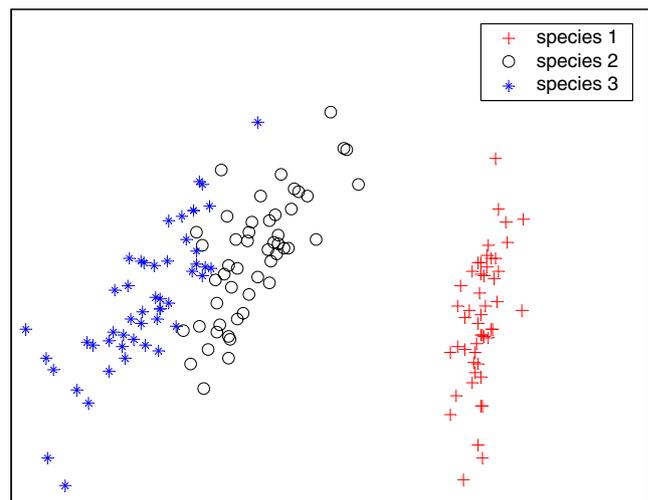


Fig. 16 Projection of two principal components of the Iris data. Species 2 and 3 are much alike from the overlapping of their respective data set

Table 4 Numbers of the encoded nodes

Iteration times		1	2	3	4	5	6	7	8	9	10
Node numbers	ART 3	4	5	6	7	8	9	10	11	12	13
in memory	ReART	3	3	3	3	3	3	3	3	3	3

When iterating once more, ART 3 adds one more node, but ReART holds only 3 nodes stably to store the input patterns.

examples. We choose digits “1,” “6,” and “9,” 200 samples for each digit image, totally 600 input patterns. Samples are shown in Fig. 14.

First, decrease the input vectors’ dimension from 784 down to 2 by locally linear embedding (LLE) (Roweis and Saul 2000). Secondly, use ART 3 and ReART to do the clustering. The recognition rate in up to 98%, and Fig. 15 shows the time consumption. Obviously, ReART runs faster than ART 3. Owing to the introduction of NO retrograde, after each reset wave, the winner node loses more transmitters, which reduces the time consumption.

3.2.3 Time complexity analysis

The time complexity can be analyzed as follows: presynaptic transmitter u_{ij} is released at a rate jointly proportional to S_i and a function $f(x_j)$, of the postsynaptic activity. Assume that at time $t=0$, transmitter u_{ij} has accumulated to its maximal level z_{ij} and that activity x_j and bound transmitter v_{ij} equal 0. Consider a time interval $t=0^+$ immediately after S_i arrives at the synapse. The amount of bound transmitter is determined by equation:

$$\frac{dv_{ij}}{dt} = -v_{ij} + u_{ij}S_i f(x_j) - v_{ij}[\text{reset signal } \zeta]. \tag{18}$$

At time $t=0^+$,

$$\frac{dv_{ij}}{dt} \approx z_{ij}S_i f(0). \tag{19}$$

In the next time interval, x_j has grown up to a positive value. Then as a retrograde messenger, NO diffuses out of

postsynaptic sites into presynaptic sites, enhancing the transmitter releasing by Eq. (11):

$$\frac{du_{ij}}{dt} = (z_{ij} - u_{ij}) - u_{ij}\xi + u_{ij}x_j$$

So the reset wave can decrease the mismatch weights’ functional value u_{ij} much more than ART 3, and ReART causes less mismatch nodes winning before the right category node is selected, and can converge faster. But ART 3 may spend much more time on inactivating the mismatch nodes.

3.3 Comparison of memory weights on Fisher’s Iris data

Fisher’s Iris data was first used by Ronald Fisher (Fisher 1936). There are four discriminating variables (petal, sepal length, petal and petal width) and three species of iris (Setosa, Virginica and Versicolor). It consists of 50 specimens for each of the species. The features are shown in Table 3 (Quah et al. 2005). Figure 16 gives the principal component analysis of Iris data (Jolliffe 1986; Ahn and Oh 2003). We can see that the Virginica and Versicolor classes are much alike from the overlapping of their respective data sets. In this part, we use Iris data to verifying the advantages of the forgetting mechanism.

Design the learning rate parameter d becoming smaller and smaller as time goes on. One hundred fifty patterns are presented repeatedly till the clustering is stable. When $\rho=0.98$, the recognition rate of ART 3 and ReART are both 98.67%. The node numbers in memory after each iteration are illustrated in Table 4.

Table 5 When iterating thirdly, the top-down LTM weights in memory

Nodes	ART 3	Nodes	ReART
1	0.6286 0.4263 0.1743 0.0000	1	0.6286 0.4263 0.1743 0.0000
2	0.7622 0.3484 0.5528 0.1632	2	0.7482 0.3423 0.5402 0.1600
3	0.6421 0.2932 0.4305 0.1263	3	0.8390 0.3709 0.7064 0.2492
4	0.7633 0.3493 0.5554 0.1646		–
5	0.7633 0.3493 0.5554 0.1646		–
6	0.8390 0.3709 0.7064 0.2492		–

ART 3 encodes nodes 2, 4 and 5 with the similar weights. Comparatively, ReART encodes only 3 nodes to store the input patterns.

Table 5 shows the top–down LTM weights in memory when iterating thirdly. ART 3 encodes nodes 2, 4 and 5 with the similar weights. Comparatively, ReART encodes only three nodes to store the input patterns.

Because ART 3 neural network is unsupervised in its learning process, the weights in memory will be modified repeatedly, causing that they deviate from the initial pattern, which induces the establishment of them. When input sequences are presented repeatedly, the former weights may not be activated because they cannot match the initial inputs by the vigilance parameter ρ , and a new node will be established in respond to the current input. The phenomenon of pattern excursion results in an increase in node numbers. By incorporating the matched nodes and abandoning the erroneous nodes according to the forgetting mechanism, ReART obtains efficient encoding in memory.

4 Discussions

Regarding data clustering, adaptive resonance theory has become an indispensable tool in many areas of engineering and attracts more and more attention in other fields. ART 3, a member of the ART neural networks family, embeds the computational properties of chemical synapses to fulfill its search process. However, ART 3 still has limitations in recognition rate, convergence speed, and weights updating mechanism.

Actually, the transmitter release process between synapses is somewhat complicated. Recent research has shown that NO, as a very small and nonpolar molecule, will spread in three dimensions away from a site of synthesis regardless of intervening cellular or membrane structures (Lancaster 1994; Wood and Garthwaite 1994). NO has strong neuro-modulating ability. Especially, in long-term potentiation, it diffuses out of postsynaptic sites into presynaptic sites, serving as a retrograde messenger. NO leads to the increase of neuro-transmitter release indirectly and guarantee the information transfer between synapse. NO not only plays an important role in neuroscience, but also gives inspiration to neural networks and other neural computation.

Based on this motivation, this paper presents retrograde ART (ReART), an improved adaptive resonance model embedding NO retrograde mechanism. We introduce NO retrograde mechanism into ART 3 model and establish feedback in the search–reset cycle. To be more specific, our contributions can be listed as following:

First, a novel efficient search hypothesis is proposed. By investigating the different ways of measuring the distance between two vectors, we check the input vector and the expectation vector element by element to obtain the

distance information. The mismatched elements are reduced to zero according to the novel transmitter released hypothesis. The simulation results on some analog samples show that our ReART model gets better categories than the original ART 3. However, the rigorous mathematical proof remains unsolved.

Second, the convergence speed of an algorithm is a key issue. In ART 3, after a sample vector is input, the search for a proper node to be activated sometimes sticks into a search–reset–search circle and converges slowly at the right node. In this paper, the presynaptic transmitter accumulation and releasing mechanism is revised, leading to stable and fast convergence of the searching process. Based on NO retrograde in LTP, a novel transmitter released mechanism is presented. As a retrograde messenger, NO diffuses to the presynaptic synapse from the postsynaptic synapse. It enhances the amounts of presynaptic transmitter to be released, resulting in an increase of the neurotransmitter received by the postsynaptic node, as shown in Fig. 7. Mathematically, the improved release equation is formulated. In application to data classification, ReART converges faster. But it should be noted that the convergence problem of the novel search mechanism has not been thoroughly analyzed and proved yet, and further study is still necessary.

Third, another important function of human brain is oblivion. To avoid the potential phenomenon of pattern excursion during the weights learning process, the forgetting mechanism is introduced to ReART. By incorporating the matched nodes and abandoning the erroneous nodes, the proposed model regulates the LTM weights dynamically, and so efficiently makes use of the memory space. Simulation results demonstrate that to store the same samples, ReART uses fewer nodes than the original ART 3 and reduces memory consumption.

We human beings can rapidly learn new environments without rapidly forgetting what we have already known. As a result, we don't need to fear that we will suddenly forget our parents' faces when learning to recognize a new friend's face. This is sometimes called the problem of catastrophic forgetting. The problem that the brain learns quickly and stably without catastrophic forgetting its past knowledge has been called the stability–plasticity dilemma (Grossberg 1980). Every brain system needs to solve the dilemma. In the ART neural networks, if a good enough match occurs between bottom–up signal and learned top–down expectation, previously learned knowledge can be refined. If new samples can't form a good enough match with the LTM weights stored in memory, new category nodes will be encoded instead of catastrophic forgetting of an old one. But it requires that the network has enough storage space.

Table 6 The LTM weights of Fisher's Iris data by different vigilances

Vigilance	LTM weights			
	Nodes	ART 3	Nodes	ReART
$\rho=0.98$	1	[0.6286 0.4263 0.1743 0.0000]	1	[0.6286 0.4263 0.1743 0.0000]
	2	[0.7622 0.3484 0.5528 0.1632]	2	[0.7482 0.3423 0.5402 0.1600]
	3	[0.6421 0.2932 0.4305 0.1263]	3	[0.8390 0.3709 0.7064 0.2492]
	4	[0.7633 0.3493 0.5554 0.1646]		–
	5	[0.7633 0.3493 0.5554 0.1646]		–
	6	[0.8390 0.3709 0.7064 0.2492]		–
$\rho=0.97$	1	[0.6288 0.4243 0.1776 0.0021]	1	[0.6288 0.4243 0.1776 0.0021]
	2	[0.6383 0.2931 0.4581 0.1374]	2	[0.7511 0.3434 0.5437 0.1606]
	3	[0.8389 0.3705 0.7053 0.2484]	3	[0.8390 0.3709 0.7064 0.2492]
	4	[0.7661 0.3501 0.5552 0.1638]		–
$\rho=0.96$	1	[0.6300 0.4184 0.1891 0.0075]	1	[0.6300 0.4184 0.1891 0.0075]
	2	[0.7092 0.3150 0.5033 0.1425]	2	[0.7608 0.3482 0.5641 0.1711]
	3	[0.8195 0.3645 0.6655 0.2239]	3	[0.8464 0.3698 0.7068 0.2469]
	4	[0.7833 0.3679 0.5674 0.1748]		–

When the vigilance varies, the number of LTM weights in ART 3 changes. On the contrary, thanks to the NO retrograde and the forgetting mechanism, ReART categorizes stably and quickly.

It's important to point out some limitations of ReART still remain, and future work should be focused on:

- 1) Selection of the vigilance parameter ρ . In most unsupervised ART systems, when the networks receive input samples, it needs to test for some times to choose the right value of ρ . Low vigilance allows broad generalization, coarse categories and abstract memories. High vigilance leads to narrow generalization, fine categories and detailed memories (Carpenter and Grossberg 2002). The LTM weights of Fisher's Iris data (as shown in Fig. 16) with different vigilances are shown in Table 6. It is obviously seen that our ReART gives better results than ART 3 with various vigilances, which means the introduction of NO retrograde and forgetting mechanism make ReART robust to the selection of the vigilance parameter ρ . However, the vigilance value is still limited in a small region and should be decided by experience. Some supervised learning algorithms have been presented to overcome this shortcoming (Stephen et al. 2005), but research on an adaptive vigilance to unsupervised algorithms is still needed.
- 2) Sensitive to noise vectors. The ART neural networks think that every input vector is correct and dependable. When the network finds a matched node to the current sample X , the corresponding LTM weights W_i will be modified based on the current input vector.

$$W_i = W_i + (X - W_i) \quad (20)$$

To stabilize the weights modifying process, another equation is given:

$$W_i = W_i + d(X - W_i) \quad (21)$$

where d represents the learning rate, which is decreasing as time goes on. If X is a noise vector, the network still modifies the matched LTM weights. So it is sensitive to noise. Although, the forgetting mechanism introduced in this paper can reduce noise-induced memory consumption, ReART still can't recognize and avoid noise.

- 3) LTP is only one of the two major types of synaptic plasticity. Synaptic plasticity is believed to serve as a cellular model for learning and memory. It is one of the fields that progresses rapidly and has a lot of success in neuroscience. Another major type of synaptic plasticity is long-term depression (LTD). NO serves as a kind of retrograde messenger in LTD, as well as in LTP (Bolshakov and Siegelbaum 1994). Introducing LTP and LTD cooperation mechanism to ART 3 will provide new inspiration to find a novel effective classification and clustering algorithm.
- 4) Preliminary analysis on NO serving as a retrograde messenger is provided in this paper, and there are more functions of NO in neural activities. It is noteworthy to investigate the effect of NO in neural activities further and extend the method to the cases of other neural network models in future work, such as fuzzy ART (Carpenter et al. 1991b) and ARTMAP (Carpenter et al. 1991a).

Acknowledgements The authors would like to thank the action editor and the anonymous reviewer for their constructive comments. This work was supported in part by Natural Science Foundation of China (No. 60171003, 30370416, 60575044, 60675005), National Distinguished Young Scholars Fund of China (No. 60225015).

References

- Ahn, J.-H., & Oh, J.-H. (2003). A constrained EM algorithm for principal component analysis. *Neural Computation*, *15*, 57–65.
- Bolshakov, V. Y., & Siegelbaum, S. A. (1994). Postsynaptic induction and presynaptic expression of hippocampal long-term depression. *Science*, *264*, 1148–1152.
- Bredt, D., & Snyder, S. (1992). Nitric oxide, a novel neuronal messenger. *Neuron*, *8*, 3–11.
- Burette, A., Zabel, U., Weinberg, R. J., Schmidt, H. H. W., & Valtchanoff, J. G. (2002). Synaptic localization of nitric oxide synthase and soluble guanylyl cyclase in the hippocampus. *Journal of Neuroscience*, *22*, 8961–8970.
- Carpenter, G. A., & Grossberg, S. (1987). 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. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, *3*, 129–152.
- Carpenter, G. A., & Grossberg, S. (2002). Adaptive resonance theory. *The Handbook of Brain Theory and Neural networks, second edition*.
- Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1991a). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, *4*, 565–588.
- Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, *4*, 759–771.
- Chang, H. C., Kopaska-Merkel, D. C., Chen, H. C., & Durrans, S. R. (2000). Lithofacies identification using multiple adaptive resonance theory neural networks and group decision expert system. *Computers & Geosciences*, *26*, 591–601.
- Christelle, L. M. B., & Garthwaite, J. (2003). On the role of nitric oxide in hippocampal long-term potentiation. *Journal of Neuroscience*, *23*(5);: 1941–1948.
- Feng, R., & José, L. C. (2006). Magnetoencephalographic artifact identification and automatic removal based on independent component analysis and categorization approaches. *Journal of Neuroscience Methods*, *157*, 337–354.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179–188.
- Garthwaite, J., Charles, S., & Chess-Williams, R. (1988). Endothelium-derived relaxing factor release on activation of NMDA receptors suggests role as interneurons messenger in the brain. *Nature*, *336*, 385–388.
- Gladys, Y. K., & Kelly, T. P. (1999). Nitric oxide acts as a postsynaptic signaling molecule in calcium/calmodulin-induced synaptic potentiation in hippocampal CA1 pyramidal neurons. *Journal of Neuroscience*, *19*, 6784–6794.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, *87*, 1–51.
- Grossberg, S., Govindarajan, K. K., Wyse, L. L., & Cohen, M. A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. *Neural Networks*, *17*, 511–536.
- Jolliffe, I. (1986). *Principal component analysis*. Berlin Heidelberg New York: Springer.
- José, L. C., & Wolfram, S. (1999). A predictive reinforcement model of dopamine neurons for learning approach behavior. *Journal of Computational Neuroscience*, *6*, 191–214.
- Kohonen, T. (1997). *Self-organizing maps*. (2nd ed.). Berlin Heidelberg New York: Springer.
- Kumar, S., & Guez, A. (1991). ART based adaptive pole placement for neural controllers. *Neural Networks*, *4*, 319–335.
- Lancaster, J. R. (1994). Simulation of the diffusion and reaction of endogenously produced nitric oxide. *Proceedings of the National Academy of Sciences USA*, *91*, 8137–8141.
- Ledo, A., Barbosa, R. M., Gerhardt, G. A., Cadenas, E., & Laranjinha, J. (2005). Concentration dynamics of nitric oxide in rat hippocampal subregions evoked by stimulation of the NMDA glutamate receptor. *Journal of Neuroscience*, *102*, 17483–17488.
- Levine, D. S., & Penz, P. A. (1990). ART 1.5-A simplified adaptive resonance network for classifying low-dimensional analog data. In *International Joint Conference on Neural Networks (Washington)*, (pp. 639–642). Hillsdale, NJ: Erlbaum.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, *1* (pp. 281–297). Berkeley, CA: University of California Press.
- Malenka, C. R., & Nicoll, A. R. (1999). Long-term potentiation—A decade of progress? *Science*, *285*.
- Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. E. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, *14*, 1215–1229.
- Philippides, A., Husbands, P., & O’Shea, M. (2000). Four-dimensional neuronal signaling by nitric oxide: A computational analysis. *Journal of Neuroscience*, *20*, 1199–1207.
- Quah, K. H., Quek, C., & Leedham, G. (2005). Reinforcement learning combined with a fuzzy adaptive learning control network (FALCON-R) for pattern classification. *Pattern Recognition*, *38*, 513–526.
- Rosenberg, P. A., Li, Y., Le, M., & Zhang, Y. (2000). Nitric oxide-stimulated increase in extracellular adenosine accumulation in rat forebrain neurons in culture is associated with ATP hydrolysis and inhibition of adenosine kinase activity. *Journal of Neuroscience*, *20*, 6294–6301.
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*, 2323–2326.
- Smith, T. M. C., & Philippides, A. (2000). Nitric oxide signalling in real and artificial neural networks. *British Telecom Technology Journal*, *18*, 140–149.
- Solan, Z., & Ruppin, E. (2001). Similarity in perception: A window to brain organization. *Journal of Cognitive Neuroscience*, *13*, 18–30.
- Stephen, J. V., Gregory, L. H., & Michael, G. (2005). Boosted ARTMAP: Modifications to fuzzy ARTMAP motivated by boosting theory. *Neural Networks*, *1*–23.
- Whitlock, J. R., Heynen, A. J., Shuler, M. G., & Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *Science*, *313*, 1093–1097.
- Williamson, J. R. (1997). A constructive, incremental-learning network for mixture modeling and classification. *Neural Computation*, *9*, 1517–1543.
- Wood, J., & Garthwaite, J. (1994). Models of the diffusional spread of nitric oxide: Implications for neural nitric oxide signaling and its pharmacological properties. *Neuropharmacology*, *33*, 1235–1244.
- Yang, T. (2003). Multi-layer cellular neural networks: Theory and applications to modeling nitric oxide diffusion in nervous systems. *International Journal of Computational Cognition*, *1*, 1–23.
- Yoshihara, M., Adolfsen, B., Galle, K. T., & Littleton, J. T. (2005). Retrograde signaling by Syt 4 induces presynaptic release and synapse-specific growth. *Science*, *310*, 858–863.