Deep Adaptive Resonance Theory for Learning Biologically Inspired Episodic Memory

Gyeong-Moon Park
School of Electrical Engineering
KAIST
Daejeon, Republic of Korea
Email: gmpark@rit.kaist.ac.kr

Jong-Hwan Kim, Fellow, IEEE
School of Electrical Engineering
KAIST
Daejeon, Republic of Korea
Email: johkim@rit.kaist.ac.kr

Abstract—Biologically inspired episodic memory is able to store time sequential events, and to recall all of them from partial information. Because of the advantages of episodic memory, the biological concepts of episodic memory have been utilized to many applications. In this research, we propose a new memory model, called Deep ART (Adaptive Resonance Theory), to make a robust memory system for learning episodic memory. Deep ART has an attribute field in the bottom layer, which is newly designed to get semantic information of inputs. After encoding all inputs with their features, events are categorized in the event field using inputs. Since an episode is made of a temporal sequence of events, Deep ART makes event sequences with proposed sequence encoding and decoding processes. They can encode any temporal sequence of events, even if there are duplicated events in the episode. Moreover, based on the result of the analysis of retrieval error, Deep ART does not use complement coding for partial inputs to enhance the accuracy of episode retrieval from partial cues. The simulation results demonstrate the effectiveness of Deep ART as the long term memory.

I. INTRODUCTION

The declarative memory is the memory capable of conscious recollection. It is mainly divided into semantic memory and episodic memory [1]. Semantic memory is the memory based on facts about the world, which can be shared and accepted between each other. On the other hand, episodic memory is a subjective memory from personal experiences. Since an episode is composed of time sequential events [2], episodic memory stores temporal sequences of events incrementally.

In general, human is able to recall related situations or behaviors from partially observed information, using episodic memory [3], [4]. It gives human lots of cognitive capabilities [5]. For instance, human can infer various situations unseen from observations and predict next situations based on his or her experiences. Moreover, these experiences can be applied to various environments or purposes.

To utilize many advantages of episodic memory, there have been many researches applying biological concepts of episodic memory. Episodic memory-driving Markov decision processes (EM-MDPs) uses the unidirectional feature of episodic memory to solve high-dimensional problems of partially observable Markov decision processes (POMDPs) [6]. It is applied to the navigation of mobile robots. This model uses a merit of episodic memory well in terms of the applicability of experienced events, but all learned episodes are limited to one purpose: the robots’ paths. If there are episodes having different purposes, it is hard to share sequential events properly. Another recent computational model for learning episodic memory, called episodic memory-ART (EM-ART), was proposed [7], [8]. Since EM-ART is based on the ART neural model which is the unsupervised neural network, the memory model can learn new events and episodes incrementally. The event sequences are encoded by using the geometric sequence between events, and the model can retrieve a whole sequence of events after the learning process. EM-ART memory model was applied to various researches because input channels of EM-ART can be user-defined for their purposes. EM-ART used for learning user daily activity pattern [9], and emotional states [10]. Learned episodic memories could be shared between different agents [11]. Moreover, EM-ART was applied for robot’s task intelligence [12]–[14]. However, EM-ART model cannot memorize episodes that have repeated same events [8]. Also, EM-ART may occur retrieval errors depending on the numbers of events in memorized episodes, for partial cues [14]. These limitations are the main reasons that the model could not be used as episodic memory learning model for universal purposes.

In this paper, we propose Deep ART network which is a new learning model for biologically inspired episodic memory. It has a deeper structure to encode from low-level features like input attributes to high-level features like episodes, daily episode pattern, and even more for a longer period of time. A newly added attribute field encodes the characteristics of inputs. It enables Deep ART memory to recognize various states of inputs. Also, proposed encoding and decoding processes memorize temporal sequences of events properly. The episode field is newly designed to encode sequences of individual events. Using the new episode field, the sequence encoding and decoding processes make Deep ART learn any temporal sequence of events, even if episodes have duplicated events. In the retrieval process, Deep ART is able to retrieve a proper episode robustly from partial cues. The main reason that the retrieval error occurs in EM-ART networks is the complement coding process, due to the different numbers of events in memorized episodes, in the retrieval. [14]. To overcome this problem, we do not consider the complement coding in the retrieval process of Deep ART. It makes Deep ART robust to
partial inputs when it recalls an episode. The effectiveness of Deep ART is demonstrated using the simulation results with comparisons between our proposed Deep ART model and EM-ART memory model.

This paper is organized as follows. Section II presents the proposed Deep ART memory model with detailed descriptions about each process. Section III shows the simulation results along with discussions to demonstrate the effectiveness of the proposed model. Finally, Section IV presents conclusion.

II. DEEP ART

Proposed Deep ART is the unsupervised neural network to learn and recall episodic memory. To memorize semantic information of inputs, we provide the attribute field that gets inputs with their characteristics. It makes Deep ART model represent events more concretely.

Since an episode is a temporal sequence of events, encoding the temporal sequence of events is important for learning episodic memory. Previous models [7], [10], [14] for memorizing episodes have a limitation in making a sequence of events, when episodes have duplicated events. Deep ART can encode temporal sequences of events robustly with its new encoding process, which records the time order of each event in an episode, even if there are duplicated events in the episode. It makes Deep ART more suitable to learn temporal sequences of events than other models. Moreover, this new encoding process enables Deep ART to have a deeper structure than previous models because it can memorize temporal sequences of inputs without errors. Generally, when a neural model has a deep structure, higher layers discern more general features than lower layers. In this regard, Deep ART can be adopted as the learning model of episodes and also higher levels, such as daily activity pattern.

Humans can recall their personal episodes when they see just a part of episodes, or simply they see an object related with episodes [4]. Like humans’ episodic memory, Deep ART can also retrieve a proper episode from partial cues. In the previous research [14], there is a description about the condition that the retrieval error occurs in EM-ART model. But, proposed Deep ART is able to recall whole temporal sequences of events properly without any limitation, with its different retrieval process. Therefore, Deep ART can recall episodes from partial cues robustly as the humans’ memory system.

The overall architecture of Deep ART is shown in Fig. 1. As in the figure, Deep ART gets inputs from the bottom layer, which is called the attribute field $F_1$, and the upper layers of Deep ART encode events, episodes, and daily episode patterns. The structure of Deep ART in Fig. 1 can be extended to higher layers for memorizing weekly or monthly episode patterns. In this research, we use the structure having layers up to $F_4$ to learn episodic memory. The details of episode learning and retrieval processes in Deep ART is described below.

A. Input Encoding

In this paper, an event is represented as a verb clause. The verb clause is composed of an action, objects, and other modifiers for specifying the meaning. Therefore, to encode events concretely in Deep ART memory, we adopt a new field, called attribute field, that can encode inputs with their features.

As in Fig. 1, Deep ART has several input channels $F_0^n$ for specifying events, and the $n^{th}$ input channel also has various attribute channels $F_1^k$ that get the input vectors. The basic structure of attribute field is the same as that of fusion ART [15]. Fig. 2 shows the $n^{th}$ input channel with its attribute channels. The input encoding process is described below:

1) Complement coding: An attribute channel $nF_1^k$ in the attribute field gets an input vector $nT^k = \{nI^k_1, nI^k_2, \ldots, nI^k_a\}$, where $a$ is the number of all elements in the input vector. For the purpose of input normalization in the encoding process [16], the complement coding process makes the complement of input vector $nT^k$ and the activity vector $n\cdot x^k$ as follows:

$$n\cdot T^k = 1 - nT^k,$$
$$n\cdot x^k = \{nI^k_1, nI^k_2\}.$$

2) Code activation: Using the activity vector $n\cdot x^k$, the $j^{th}$ input node in the input channel $F_2^n$ is activated by the choice function:

$$nT_j = \sum_{i=1}^{k} n\gamma_j \frac{|n\cdot x^k \land n\cdot w_j^l|}{n\alpha_k + |n\cdot w_j^l|},$$

where $n\gamma_j$ is a contribution parameter, $n\cdot w_j^l$ is a weight vector of the $j^{th}$ input, $n\cdot x^k$ is a choice parameter. The fuzzy AND operator $\land$ is defined by $(p \land q)_i = min(p_i, q_i)$, and the norm $|.|$ is defined by $|p| = \sum_i p_i$ for vectors $p$ and $q$.

3) Code competition: Among all nodes in the $n^{th}$ input channel, the $J^{th}$ input node that has the largest activation value is selected as follows:

$$nT_j = \max \{nT_j : \text{for all } F_2^n \text{ node } j\}.$$

It is a process for selecting the most similar node among already learned nodes.

4) Template matching: The selected $J^{th}$ node is checked whether it is resonant or not. It is the criteria of determining the weight update. The resonance $n\cdot m_j^k$ is calculated by the following equation:

$$n\cdot m_j^k = \frac{|n\cdot x^k \land n\cdot w_j^k|}{|n\cdot x^k|} \geq n\cdot \rho^k$$

where $n\rho^k$ is a vigilance parameter. If any node is not resonant, an uncommitted node is generated and all elements of the initial weight vector for the node is set to one.

5) Template learning: If the selected node is resonant, the weight vector is updated by the following update rule:

$$n\cdot w_j^{k(new)} = (1 - n\cdot \beta^k)n\cdot w_j^{k(old)} + n\cdot \beta^k (n\cdot x^k \land n\cdot w_j^{k(old)})$$

where $n\beta^k$ is a learning rate. From the above processes, all inputs with their characteristics are categorized in each input channel of the input field $F_2$. 
6) **Event learning**: To encode events using all inputs in the input field $F_2$, one more fusion ART network is used for the next layer. It is called the event field $F_3$ in Fig. 1. The overall process of encoding events is the same as the process of encoding inputs. All learned events are memorized in the input channel $^iF_3$ in the event field $F_3$.

**B. Episode Learning**

The episode learning process comprises two steps: encoding a sequence of events, and encoding an episode using the event sequence. The overall process of learning episodes is described below:

1) **Time encoding process**: In this research, a new sequence encoding technique is proposed for episode learning. Fig. 3 shows the time encoding process of sequential events in the event field $F_3$. The event field $F_3$ has two more channels with the existing input channel, which are buffer and output channels. When an input event comes to the input channel $^iF_3$, the output vector $\alpha y_n$ in the output channel $\alpha F_3$ is calculated by the weighted sum of input and buffer vectors as follows:

$$b x_n = a w \alpha y_{n-1}$$

$$\alpha y_n = i w i x_n + b w b x_n$$

$$= i w i x_n + b w a w \alpha y_{n-1}$$

(6)

where $i x_n$ and $b x_n$ are input and buffer vectors, and $i w$, $b w$, and $a w$ are weights for input, buffer, and output vectors. These weights are larger than zero. In the process, $\alpha y_0$ is initialized as a zero vector, and weights $i w$, $b w$, and $a w$ are set to one, two, and one. Note that a buffer vector temporally saves a previous output vector multiplied by the output weight $a w$. When a new input event comes to the input channel, the output vector gets the information of the previous sequence from the buffer channel, and encodes a new sequence of events with consideration of a new input event, based on the previous sequence. Using (6), the output vector $\alpha y_n$ finally memorizes a whole temporal sequence of input events.

2) **Sequence normalization**: After calculating the event sequence, the output vector $\alpha y_n$ needs to be normalized before it is to be a new input for the next layer. To normalize the output vector, the maximum element of this vector is selected and the positional number of it is identified. Then the normalized...
output vector $\alpha \vec{y}_n$ is calculated by the vector $\alpha \vec{y}_n$ divided by
the positional number. If the maximum element is used to
normalize the output vector, some elements can be irrational
numbers. Since it is hard to convert these elements into the
original values, the maximum positional number is used
to normalize the output vector with preserving the rational
characteristic of the vector.

3) Episode encoding: To encode an episode in the episode
field $F_3$, the normalized output vector $\alpha \vec{y}_n$ is used as a new
input. The structure of episode encoding is the same as that of
fusion ART. Eventually, all learned episodes are categorized
in the episode field $F_3$. Furthermore, Deep ART has an
extendability because of the proposed encoding process with
a deep structure. Therefore, Deep ART is able to memorize
high-level features such as daily episode pattern and weekly
episode pattern, using the proposed encoding process.

C. Episode Retrieval

After learning episodes in the episode field $F_3$, Deep ART
memory can retrieve a specific episode. Fig. 4 shows the
retrieval process for the $j^{th}$ episode. The retrieval process
is composed of the following four steps.

1) Episode retrieval from the episode field: A key idea for
retrieving learned episodes is that an input vector is substituted
by a learned weight vector. From (5), as the learning iteration
increases, the learned weight vector converges to the input
vector. It can be used for recalling input vectors using weight
vectors. Therefore, the normalized output vector is retrieved
by the weight vector connecting the episode field and the event
field, as follows:

$$\alpha \vec{y} = Jw_S$$ (7)

2) Sequence de-normalization: In case of the decoding
process of event sequences, the retrieved output vector $\alpha \vec{y}$
should be de-normalized first. For de-normalization, the max-
imum positional number among all elements in the vector
is identified. Then, $\alpha \vec{y}$ is the vector $\alpha \vec{y}$ multiplied by
the positional number. It is the opposite process of the sequence
normalization done in the encoding process.

3) Time decoding process: The retrieved output vector $\alpha \vec{y}$
is converted into bit sequences. Then, the buffer channel gets
the vector having bit sequences. To decode each event from the
sequence, each node of the buffer channel gives one element
to the input node, one at a time. Because of the winner-take-all
strategy, the decoded input vector $i \vec{x}$ has a zero vector except
for just one element that has the value one.

4) Event retrieval: From each input vector $i \vec{x}$, the node
that has the value one is selected and retrieved by the same
rule as (7). Input channels in the input field $F_3$ are substituted
by learned weights between the event field and the input field.
Finally, attribute inputs can be recalled using the same process
above. This process is recursively performed from the top layer
to the bottom layer to retrieve the whole temporal sequences
of events.

D. Robust episode retrieval from partial cues

According to the paper [14], the retrieval error from partial
cues may be occurred by a complement of an input vector.
To overcome this limitation, Deep ART does not use the
complement coding process in the retrieval process to remove
the effect of the complement vector for input cues for reducing
retrieval errors. For analyzing the effect of the complement
coding to the retrieval errors, consider that an event related
with the $j^{th}$ episode and unrelated with the $j^{th}$ one comes
to the attribute field. For simplicity, $\alpha \approx 0$. Then, the choice
function (2) for the $j^{th}$ episode becomes:

$$T_j = \frac{|x \land w_j|}{|w_j|}$$ (8)

Since the number of channels from the input field is one, it is
for the single channel and $\gamma = 1$. In case of the denominator
of (8), the weight vector converges to the input vector, as
the iterations increase. Therefore, the norm of the $j^{th}$ weight
vector is

$$|w_j| \approx |x_j|$$

$$= \sum_{n=1}^{N_j} I_n^j + \sum_{n=1}^{N_j} (1 - I_n^j) + (N - N_j)$$

$$= \sum_{n=1}^{N_j} (1) + (N - N_j)$$

$$= N$$ (9)
derive the values of \( N \) for the purpose of removing the retrieval error, assume that \( j \) goes to one for the extreme case. Since \( I_p \), \( I_i \) and \( I_j \) can have the value one for their maximum values, we get

\[
T_j - T_i = \left( (\eta_1 + \eta_2)I_p - \eta_2 \left( \sum_{n=N_i} I_n^2 \right) \right) / N \tag{13}
\]

To make the above equation positive for all cases, \( \eta_2 \) should be zero.

In summary, the complement coding is needed to normalize input vectors in the category learning process, but it can make retrieval errors from partial cues in the retrieval process. Therefore, Deep ART uses the complement coding in the encoding process only, and it does not use complements for partial inputs in the retrieval process. The simulation results will demonstrate this analysis in the following section.

III. SIMULATION RESULTS

To evaluate the performance of our memory model, the recent memory model, EM-ART, was simulated together with Deep ART for comparisons. In this paper, we simulated three episodes: 1) wipe a table; 2) prepare juice and cereal; and 3) cook a frozen pizza. All lists of episodes and sequential events are written below:

- **Wipe a table**: Approach a washcloth on a table, Grasp the washcloth, Wipe the table with the washcloth, Release the washcloth on the table
- **Prepare juice and cereal**: Approach a juice bottle on a table, Grasp the juice bottle, Move the juice bottle to a cup, Tilt the juice bottle to the cup, Release the juice bottle on the table, Approach a milkbox on a table, Grasp the milkbox, Move the milkbox to a bowl, Tilt the milkbox...
to the bowl, Release the milkbox on the table, Approach a cereal on the table, Grasp the cereal, Move the cereal to the bowl, Tilt the cereal to the bowl, Release the cereal on the table

- Cook an instant pizza: Put an instant pizza on a plate, Grasp the plate, Approach a microwave, Open the microwave, Put the plate on the microwave, Close the microwave, Push a button on the microwave, Open the microwave, Take out the plate from the microwave, Close the microwave, Move the plate on a table

A. Retrieving episodes that have duplicated events

After learning all episodes for both memory models, each model retrieved all episodes using full-length sequential events for the inputs. Both models could retrieve “wipe a table” and “prepare juice and cereal” properly because these episodes did not have any same duplicated events. However, in case of “cook an instant pizza”, there were duplicated events such as “open the microwave” and “close the microwave”. The retrieved results were listed in Table I. From the table, EM-ART could not retrieve all sequential events properly, particularly in orders of duplicated events. In case of Deep ART, all sequential events were retrieved completely. It was important to memorize episodes that had the same events because episodes could include repeated same events in many practical cases. Furthermore, EM-ART did not have the attribute field, the model could not memorize attributes for inputs. In this paper, we made input groups that were used for EM-ART’s inputs, such as “instant_pizza”, “on_plate”, and so on. Deep ART could distinguish various states of an object, using the attribute field automatically. Also, Deep ART model could recall associated words or objects from partial states of inputs.

<table>
<thead>
<tr>
<th>Sequential events retrieved by each model</th>
<th>EM-ART</th>
<th>Deep ART</th>
</tr>
</thead>
<tbody>
<tr>
<td>Put instant_pizza on plate</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Grasp plate</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Approach microwave</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Put plate on microwave</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Push button on microwave</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Open microwave</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Take out plate from microwave</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Close microwave</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Move plate on table</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

B. Episode retrieval from partial cues

The retrieval error occurred when there were episodes that had different lengths of event sequences. Since all learned episodes in this paper had different sequences, retrieval error rates between EM-ART and Deep ART could be compared to check the retrieval performances. In fact, as you could see in the episode lists above, the situation that each episode could have a different sequence is quite common. Therefore, this simulation was valuable to determine whether the memory model could be adopted or not for real applications robustly.

In (13), the retrieval error occurred more easily when the difference between the shortest event sequence (e.g. $N_i$) and the longest event sequence (e.g. $N_j$) was large. Therefore, to evaluate the accuracy of the episode retrieval from partial cues, we made partial inputs using the episode that has the longest sequence of events. Since the length of the longest sequence of events is fifteen, the combination of all partial inputs is $2^{15} - 1 = 32767$. The shortest sequence of events was set from one to eight. The comparison results of retrieval error rates (%) for two memory models were shown in Fig. 5 and Table II.

![Fig. 5. The retrieval error rates (%) of two models according to the smallest number of events among learned episodes.](image)

<table>
<thead>
<tr>
<th>The smallest number of events</th>
<th>Retrieval error rate (%)</th>
<th>EM-ART</th>
<th>Deep ART</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>28.4982 (9338/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14.9602 (4902/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>7.5381 (2470/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3.7599 (1232/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.8769 (615/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.9369 (307/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.4639 (152/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.2319 (76/32767)</td>
<td>0 (0/32767)</td>
<td></td>
</tr>
</tbody>
</table>

As expected, the retrieval error occurred in EM-ART, when the numbers of sequential events of individual episodes were different. Moreover, the retrieval error rate exponentially decreased as the smallest number of events increased. However, Deep ART was not influenced by the difference of sequential events. It demonstrates the effectiveness of our analysis based on (13).
IV. CONCLUSION

In this paper, we proposed a new memory model, called Deep ART neural model, for learning biologically inspired episodic memory. It was based on unsupervised neural network to categorize events and episodes from inputs automatically. Since it could increase the number of nodes for each channel adaptively, it was suitable for incremental learning, like the human’s memory system. Deep ART had the attribute field in the bottom layer which could memorize inputs with their attributes together. The attribute field helped Deep ART to encode states of inputs, and to represent events concretely. Newly proposed sequence encoding and decoding processes were presented for Deep ART. The main idea of encoding and decoding processes was that our Deep ART focused on remembering the sequences of individual events, unlike other existing models that tried to memorize the relations between sequential events. The proposed processes were powerful when the same event was duplicated in an episode. Eventually, Deep ART could be extended for a deeper structure without encoding and retrieval errors.

Deep ART was robust to partial cues when it retrieved a relevant episode from partial inputs. The complement coding could affect the retrieval error rates, so that previous models could not cope with all partial inputs well. Deep ART was able to retrieve a proper event sequence from partial cues because it ignored complements in the retrieval process. The effectiveness of Deep ART was demonstrated through simulation results for three episodes.

ACKNOWLEDGMENT

This work was partly supported by Institute for Information communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. B0101-15-0551, Virtual Creatures with Digital DNA) and the Technology Innovation Program, 10045252, Development of robot task intelligence technology, supported by the Ministry of Trade, Industry, and Energy (MOTIE, Korea).

REFERENCES