1993-05

Extracted Memory from Temporal Patterns Using Adaptive Resonance and Recurrent Network

EXTRACTED MEMORY FROM TEMPORAL PATTERNS USING ADAPTIVE RESONANCE AND RECURRENT NETWORK

Osamu Araki

May, 1993

Technical Report CAS/CNS–93–042

Boston University Center for Adaptive Systems and
Department of Cognitive and Neural Systems
111 Cummington Street
Boston, MA 02215

Extracted Memory from Temporal Patterns using Adaptive Resonance and Recurrent Network

Osamu Araki[1]
Department of Cognitive and Neural Systems, Boston University
111 Cumington Street, Boston, MA 02215

Abstract

We can recognize objects through receiving continuously huge temporal information including redundancy and noise, and can memorize them. This paper proposes a neural network model which extracts pre-recognized patterns from temporally sequential patterns which include redundancy, and memorizes the patterns temporarily. This model consists of an adaptive resonance system and a recurrent time-delay network. The extraction is executed by the matching mechanism of the adaptive resonance system, and the temporal information is processed and stored by the recurrent network. Simple simulations are examined to exemplify the property of extraction.

1 Introduction

Neural networks of the brain are doing real-time and parallel processing continuously for afferent signals including visual, somatosensory, and the other sensory information. How do the neural networks work to remember such temporal patterns as a sequential scene of a movie ?

Concerning temporal pattern recognition, a lot of models have been proposed, for example, time-delay network (Waibel, 1989), recurrent network (Williams & Zipser, 1989), and spatial patterns (Grossberg, 1986). The time-delay network cannot recognize temporal patterns which have arbitrary interval time. The recurrent model supposes that all temporal patterns are to be recognized and/or stored, even if the patterns include redundancy. So, the computational cost is too much. Grossberg's model supposes that input patterns are itemized.

The proposed model in this paper, which consists of an adaptive resonance system and a recurrent network extracts and recalls pre-recognized patterns from such huge temporal patterns as include redundancy and arbitrary interval. This model provides a fundamental mechanism for the extraction in temporal information processing.

2 Extracted Memory from Temporal Patterns

The proposed model consists of the two subsystems, which are an adaptive resonance system and a recurrent network (Figure 1). The former part consisting of the field F_1 and F_2 is a main module of adaptive resonance theory (ART) proposed by Carpenter and Grossberg (1987), and the latter part consisting of F_2 and F_3 is a recurrent time-delay network.

ART classifies input patterns into a category without teacher by matched between bottom-up and top-down signals. The field F_2, which receives signals from F_1 represents a category of the input patterns. F_1 receives input pattern I and top-down expectation signal V^J from F_2. A mismatch is detected, when the vigilance parameter is $\rho > \frac{|I \cap V^J|}{|I|}$. Otherwise, resonance is said to occur. The mismatch suppresses the current active node and the next best matching node is chosen in F_2.

This paper assumes that the model has three phases, a recognition phase, a learning phase, and a recall phase. In recognition phase, the model learns some patterns by using the ART module,

[1]Supported in part by Matsushita Electric Industrial Co., Ltd., Tokyo Information Systems Research Laboratory, Tokyo, Japan.

which is called "pre-recognized" in this paper. In learning phase, the model learns the order of input patterns which are changing gradually with shorter duration. The field F_2 receives minor effects from F_3, but major effects from bottom-up F_1. And in recall phase, the model recalls the input patterns from a cue signal for F_2 without receiving signals from F_1.

Mathematical representations of each node's activity and weights between nodes are shown next by using the following notations: A, B, C, D, and E are non-negative constants, I_i is an input signal, $y_j^{(i)}$ (i=1,2,3) is an activity of jth node in F_i, each of J and K is an indicator of a winner node, $z_{kl}^{(ij)}$ (i,j=1,2,3) is a connection weight from kth node in F_i to lth node in F_j, τ is a delayed time from F_2 to F_3, and f(w) is a sigmoidal function which suppresses the small level and amplifies the middle level. τ is supposed to small enough to recognize the order of input patterns.

The node activity in F_1, which receives input patterns and top-down signals follows this shunting equation.

$$\frac{dy_i^{(1)}}{dt} = -Ay_i^{(1)} + (B - y_i^{(1)})(I_i + z_{Ji}^{(21)}y_J^{(2)}) - y_i^{(1)}\sum_{j\neq i} I_j \tag{1}$$

where A is a decay rate and B is an upper bound. The weights $z^{(12)}$ and $z^{(21)}$ only change in recognition phase following the next equations.

$$\frac{dz_{ij}^{(12)}}{dt} = \begin{cases} y_J^{(2)}(y_i^{(1)} - Cz_{ij}^{(12)}) & \text{if } j = J \\ 0 & \text{otherwise} \end{cases} \qquad \frac{dz_{ji}^{(21)}}{dt} = \begin{cases} y_J^{(2)}(y_i^{(1)} - Cz_{ji}^{(21)}) & \text{if } j = J \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{dy_j^{(2)}}{dt} = -Ay_j^{(2)} + (B - y_j^{(2)})[f(y_j^{(2)}) + D\sum_i z_{ij}^{(12)}y_i^{(1)} + E\sum_k z_{kj}^{(32)}y_k^{(3)}(t)] - y_j^{(2)}\sum_{k\neq j} f(y_k^{(2)}) \tag{2}$$

where $D = 1.0\&E = 0$ in recognition phase, $D = 1.0\&E = 0.1$ in learning phase, and $D = 0.1\&E = 2.0$ in recall phase. The field F_3 receives delayed signals from F_2 as follows:

$$\frac{dy_k^{(3)}}{dt} = -Ay_k^{(3)} + (B - y_k^{(3)})f(y_k^{(2)}(t-\tau)) - y_k^{(3)}\sum_{m\neq k} f(y_m^{(3)}) \tag{3}$$

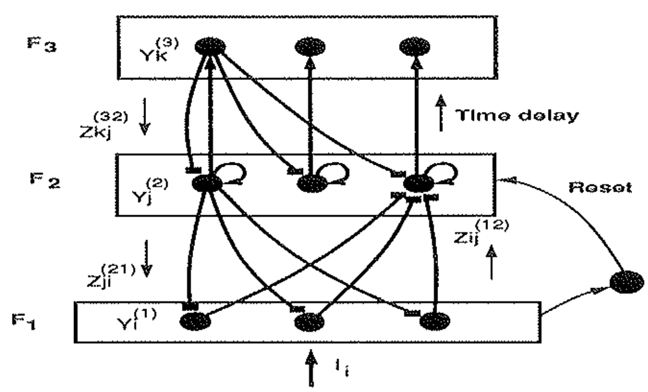


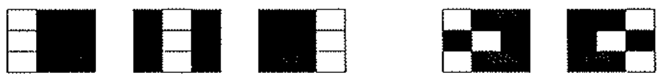
Figure 1: The architecture of the proposed model.

The weight $z^{(32)}$ only changes in learning phase, following this.

$$\frac{dz_{kj}^{(32)}}{dt} = \begin{cases} y_K^{(3)}(y_j^{(2)} - Cz_{kj}^{(32)}) & \text{if } k = K \\ 0 & \text{otherwise} \end{cases}$$

The node activity $y_j^{(2)}$ receives excitatory signals from F_1 and F_3. The signals from F_3 representing delayed information in F_2 are returned back to F_2 again so that the model can wait for long time until another pre-recognized pattern is input. When the pattern is input, $z^{(32)}$ learns the category in F_2 quickly. Thus, this model memorizes the temporal order of the shown patterns, but it cannot memorize them all, because resonance to unknown patterns takes longer time to develop than resonance to known patterns.

3 Simulation



(a) (b) (c) (d) (e)

Figure 2: The recognized input patterns. The left three patterns are for the first simulation, and all five patterns are for the second one.

In recognition phase, each input pattern should be given to this system for long enough until $z^{(12)}$ and $z^{(21)}$ are stable (Figure 2). The patterns (a), (b), (c) in Figure 2 are pre-recognized in the first simulation, and the rest two patterns are added in the second simulation. For simplicity of analysis, the input patterns to this model are supposed to have small size (3×3), and in the figures shown below, the lightness represents the activities for each cell, for example, the white means 1 and the black means 0. In learning phase, gradually changing patterns are given (Figure 3). The input I_i is set to all zero, and a cue signal is given to the node in F_2 corresponding to the first ordered node in recall phase. To recall memorized patterns, cue patterns can be also given to F_1. Let the initial values for $z^{(12)}$ be positive random values smaller than 0.3. The parameters used in this simulation are as follows: $A = 1.0; B = 1.0; C = 1.0; D = 1.0; \tau = 5.0$.

t=0-5 t=5-10 t=10-15 t=15-20 t=20-25 t=25-30 t=30-35 t=35-40

Figure 3: The input patterns to F_1 which are regarded as changing temporally.

t=4 t=8 t=12 t=16 t=20

Figure 4: The recalled patterns in F_1 at each 4 unit time. The three patterns are pre-recognized.

t=4 t=8 t=12 t=16 t=20 t=24 t=28 t=32 t=36 t=40

Figure 5: The recalled patterns in F_1 at each 4 unit time. The five patterns are pre-recognized.

Figure 4 shows that the recalled pattern at each 4 unit time in the first simulation is shown in Figure 4. The vigilance parameter is $\rho = 0.7$. The values represented in Figure 4 and 5 are relative ratios of each node activity to a maximal recalled activity. The input patterns from t=0 to t=10 in Figure 3 were classified into a category of (a) in Figure 2, those of t=10-35 were classified into (b), and the rest was classified into (c). The period for each recall is longer than $\tau = 5$, because it takes at least τ to transfer information from F_2 to F_3. The pre-recognized three patterns were picked up from the input patterns and recalled sequentially as they happened. Because the vigilance parameter was low, all of the input patterns were classified into one of the three pre-recognized patterns by ART module. Though most of the input patterns were classified into a category of (b), the duration of recall (b) is as long as the others.

In the second simulation, five pre-recognized patterns were input with higher vigilance parameter ($\rho = 0.9$), so the order of finer patterns were recalled (Figure 5). The input patterns of t=0-5 are classified into (a) in Figure 2, those of t=5-10 are (d), those of t=10-20 are (b), those of t=20-25 are classified into a new category, those of t=25-30 are (e), and those of t=30-40 are (c). The pattern t=20-25 in Figure 3 were not recalled even though the pattern was classified into a new category . Because a winner node representing the new category in F_2 did not grow enough during t=20-25, the past information was retained by the self-excitatory connections within F_2 and a feedback loop from F_3 to F_2.

These simulations show that this model extracts only the pre-recognized patterns among input patterns regardless of a vigilance parameter, and recalls temporarily the order in a shorter period than they actually happened.

4 Conclusion

This model extracts pre-recognized patterns from temporally sequential patterns, and memorizes the order and the patterns except unknown patterns and the duration time of each pattern. This model provides a fundamental real-time mechanism for such a useful extraction from huge data such as temporal information processing.

On the other hand, this model cannot recognize and distinguish different input sequences. Further studies on temporal recognition, learning of unknown patterns, and learning of duration would be needed for modeling a human memory system.

5 Acknowledgement

I wish to thank Prof. Gail A. Carpenter and Dr. Gary Bradski for helpful comments.

References

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.

Grossberg,S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E.C.Schwab & H.C. Nusbaum (Eds.) *Pattern Recognition by Humans and Machines, Vol.1: Speech Perception*, Academic Press.

Waibel,A. (1989). Modular construction of time-delay neural networks for speech recognition. *Neural Computation*, **1**, 39-46.

Williams,R.J. & Zipser,D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, **1**, 270-280.