Estimation of pattern classification vigilance parameter using neural network

손 준 혁*, 서 보 혁**
(Jun-Hyug Son, Bo-Hyeok Seo)

Abstract - This paper estimates Adaptive Resonance Theory 1(ART1) as a vigilance parameter of pattern clustering algorithm. Inherent characteristics of the model are analyzed. In particular the vigilance parameter ρ and its role in classification of patterns is examined. Our estimates show that the vigilance parameter as designed originally does not necessarily increase the number of categories with its value but can decrease also. This is against the claim of solving the stability-plasticity dilemma. However, we have proposed a modified vigilance parameter estimate criterion which takes into account the problem of subset and superset patterns and stably categorizes arbitrarily many input patterns in one list presentation when the vigilance parameter is closer to one.

Key ward : neural, vigilance parameter, estimate, ART1

1. Introduction

Learning pattern was the analysis of the instability of the feed-forward instar-outstar model that led to the discovery of Adaptive Resonance theory(ART) and to development of the Neural Network(NN) systems ART1, ART2 and ART3. The instability of instar-outstar networks could be solved by reducing the learning rate gradually to zero, thus freezing the learning categories. But then the network would lose its plasticity or the ability to react to any new data. It is obviously difficult to have both stability and plasticity. ART network are designed, in particular, to resolve the stability-plasticity dilemma; that is, they are stable enough to preserve significant past learning but nevertheless remain adaptable enough to incorporate new information(clusters) whenever it might appear[1].

2. ART Theory

The theoretical background of the ART model consists of a body of nonlinear mathematics to describe the constantly changing, nonlinear behavior of psychological and psycho-physiological phenomena. Detailed information on the subject can be found in [2-7]. The central concept of the ART network is adaptive resonance. Adaptive resonance is a process that occurs through the competition of feedforward and feedback networks. A typical ART network consists of several layers of computational units or nodes across which activity patterns are generated. Since these patterns can change significantly as each stimulus is presented to the network, they constitute a type of short-term memory (STM). The STM layers are connected by slowly-varying weighted pathways, which constitute a type of long-term memory (LTM). The input STM layer receives information from the network stimulus, which is then passed through feedforward or bottom-up pathways to a template STM layer. The template STM layer forms a prototype or template category based on the previous experience stored in the LTM weights. This template is then passed back to the input STM layer through feedback or top-down pathways. The competition between the original input stimulus and the top-down template at the input STM tends to attenuate the total network activity when the STM patterns are different and tends to amplify the total network activity when the STM patterns are similar. The amplification of network activity, or resonance of matching of input and template STM patterns, is capable of stimulating a slow adaptive process that modifies the LTM weights in the feedforward and feedback pathways. The adaptation of LTM by resonant STM patterns is the adaptive resonance process. The ability of resonance conditions in the ART network to cause modification of the LTM weights is mediated by a single parameter called the vigilance parameter, which sets a resonance threshold. If the resonant condition (or lack thereof) is insufficient to pass the threshold set by the vigilance parameter, the existing template pattern is removed so that a new template pattern can be formed. This process continues until a template pattern is generated which will bring about resonant conditions that exceed the resonance threshold. Once this occurs, the LTM adaptive process is activated. The category output of the network is based on the interpretation of activity across the template layer. Typically, the most active node in the output layer is denoted as the category to which a corresponding input pattern belongs. In this paper we look at the entire output activity pattern to obtain additional category information.

The adaptive resonance process was developed in part to address the issue of stability vs. adaptability (plasticity) in a net-work. The problem deals with how an adaptive mechanism can be unresponsive to fluctuations in its environment that should not affect the desired behavior of the network while having the capability to respond quickly to environmental stimuli that are important to developing the desired behavior. In a learning sense this problem concerns the ability of a network to continually adapt and store new information it receives without losing previously stored information or storing undesirable noise that may be present in the information.

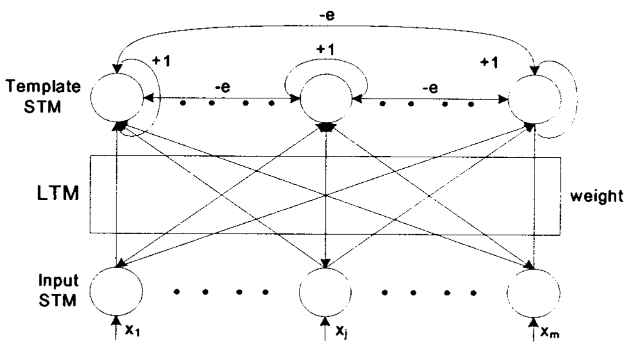

Fig. 1. ART model.

3. ART network

ART networks overcome the stability-plasticity dilemma by accepting and adapting the stored prototype of a category only when the input is sufficiently similar to it. The input and stored prototype are said to be resonate when they are sufficiently similar. When an input pattern is not sufficiently similar to any existing prototype, a new node is then created to represent a new category with the input patterns as the prototype. The meaning of sufficiently similar depends on a vigilance parameter ρ, with $0 < \rho \le 1$. If ρ is small, the similarity condition is easier to meet, resulting in a coarse categorization. On the other hand, if ρ is chosen to be close to 1, many finely divided categories are formed. The vigilance parameter value can be changed during learning such that increasing it can prompt subdivision of existing categories.

Since ART networks are for cluster discovery instead of data mapping, the above concept for resolving the stability-plasticity dilemma can be implemented by using an instar-outstar network with the desired output set equal to input. In this way, for an input pattern x, the instar part of the instar-outstar model determines the closest category that x belongs to, and then outstar part checks the similarity between the prototype of the selected category and the onput x to see if they will resonate. This is equivalent to folding the feed-forward three-layer instar-outstar network back on itself, identifying the output layer and the input layer. Hence, the competitive layer becomes the output layer and its nonzero output indicates the category a given input lies in. Thus, the minimal ART module, ART1, includes bottom-up competitive learning system combined with a top-down output pattern-learning system. ART1 is designed for binary 0/1 input, whereas ART2 is for continuous-valued inputs.

A schematic representation of ART1 is shown in Fig. 1. Each input vector x has m binary 0/1 components. Let the weights on the bottom-up links, x_j to y_i, be denoted by $\overline{w_{ij}}$ and the weights on the top-down links, y_i to x_j, be denoted by w_{ij}. Note that the first subscript of a top-down weigh indicates the source node and that second subscript indicates the destination node. Then the weight vector $w_i = (w_{i1}, w_{i2}, \cdots\cdots, w_{im})^T$, $i = 1, 2, \cdots\cdots, n$, represent stored prototype vectors and thus are also binary 0/1 vectors, where i indexes the output nodes of categories, each of which can be enabled or disabled.

3.1 Algorithm of ART network for estimation

This algorithm discovers clusters of a set of pattern vectors.

Input : A set of pattern vector x to be clustered, where $x \in \{0, 1\}^m$.

Output : A set of weight vector

$$w_i = (w_{i1}, w_{i2}, \cdots \cdots, w_{im})^T, \quad i = 1, 2, \cdots \cdots, n,$$

representing the prototype vectors of the discovered clusters, where n is the number of clusters found.

Step 0 : Initialization Set $w_{ij}(0) = 1$,

$$\overline{w}_{ij}(0) = \frac{1}{(1+m)}, \text{ for } 0 < \rho \le 1.) \tag{1}$$

Step 1 : Present a new pattern x to the input nodes.

Step 2 : Enable all the output nodes.

Step 3 : Use bottom-up processing to obtain a weighted

$$\text{sum } y_i = (\overline{w_i})^T x = \sum_{j=1}^{m} \overline{w}_{ij} x_j \tag{2}$$

where $\overline{w_{ij}}$ is the normalization of w_{ij} given by

$$\overline{w_{ij}} = \frac{w_{ij}}{\epsilon + \sum_j w_{ij}}, \; j = 1, 2, \cdots \cdots, m. \tag{3}$$

The small number ϵ(usually ϵ=0.5) is included to break ties, selecting the longer of two w_i that both have all their bits in x.

Step 4 : Use the MAX_NET procedure to find the output node i with the largest y_i value.

Step 5 : Verify that x truly belongs to the ith cluster by performing top-down processing and form the weighted sum $\sum_j w_{ij} x_j$.

Then perform the following checking :

$$\text{IF } r = \frac{\sum_{j=1}^{m} w_{ij} x_j}{|x|} > \rho, \text{ where } \|x\| = \sum_{j=1}^{m} |x_j| \quad (4)$$

THEN x belongs to the ith cluster; proceed to step 6;

ELSE IF the top layer has more than a single enabled node left, then go to step 7;

ELSE create a new output node i with its initial weights set as in step 0 and go to step 6.

Step 6 : Update the weights as follows:

$$w_{ij}(t+1) = w_{ij}(t) x_j, \ j = 1, 2, \cdots \cdots, m \quad (5)$$

which updates the weights of the ith cluster(newly created or the existing one). Then go to step 1.

Step 7 : The output node i is disabled by clamping y_i to 0. Thus, this node dose not participate in the current cluster search. The algorithm goes back to step 3 and it will attempt to establish a new cluster different from i for the pattern x. END ART1[8].

4. Estimation of vigilance parameter

An ART1 network is used to eight input patterns as shown in Fig. 2. Each input pattern is size 5*5 mask with the black(or white) grid indicating 1(or 0) and thus there are 25 nodes in the input layer($m = 25$).

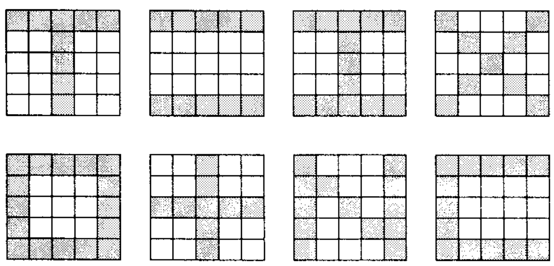

Fig. 2. Input patterns

Table 1. Comparison vigilance parameter about each number of categories

number of categories	calculated vigilance parameter	estimated vigilance parameter
2	0.356	0.346
4	0.484	0.478
6	0.723	0.712

In view of the results table 1 bear some resemblance between calculated and estimated vigilance parameter. The estimate of vigilance parameter selection have been explored in terms of both category classification and the number of winning nodes within the category pattern.

5. Conclusion

In this paper, we estimated vigilance parameter about eight input pattern with a ART network algorithm. The algorithm is based on ART network of neural network estimate to vigilance parameter. In other to estimated the input pattern of these network, we simulated pattern classification using the ART1-algorithm estimate.

Reference

[1] M. P. Davenport, A. H. Titus, "Multi-level Category Structure in the ART-2 Network", Neural Networks, IEEE Trans. on, Vol. 15, Issue 1, pp. 145 - 158, 2004.

[2] S. Grossberg, "On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems," Biol. Cybern., vol. 21, pp. 145-159, 1976.

[3] S. Grossberg, "Adaptive pattern classification and universal recoding I: parallel development and coding of neural feature detectors," Biol. Cybern., vol. 23, pp. 121-134, 1976.

[4] S. Grossberg, "Adaptive pattern classification and universal recoding II: Feed-back, expectation, olfacation, and illusions," Biol. Cybern., vol. 23, pp. 187-202, 1976.

[5] S. Grossberg, "How does a brain build a cognitive code," Psychol. Rev., vol. 87, pp. 1-51, 1980.

[6] S. Grossberg, Neural Networks for Vision and Image Processing. Cambridge, MA: MIT Press, 1992, p. 467.

[7] S. Grossberg, Pattern Recognition by Self-Organizing Neural Net-works. Cambridge, MA: MIT Press, 1991, p. 691.

[8] Chin-Teng Lin, C. S. George Lee, "Neural Fuzzy Systems-a neuro-fuzzy synergism to intelligent systems", prentice Hall P T R, 1996.