Adaptive Pointing Theory (APT) Artificial Neural Network

Kamal R. Al-Rawi and Consuelo Gonzalo

***Abstract*—The choice value and the testing process against the vigilance parameter ρ, characteristic of ART Neural Network, are merged. Only, a single unique test is required to determine if a committed category node can represent the current input or not. Advantages of APT over ART are: 1-Avoid testing every committed category node before deciding to train a committed category node or a new node must be committed, 2-The vigilance parameter ρ is fixed during training, and 3-The choice value parameter α is eliminated.**

***Index Terms*—Adaptive pointing theory, APT ANN, adaptive resonance theory, ART ANN, ARTMAP, compact fuzzy ART, artificial neural networks.**

I. Introduction

ART Artificial Neural Network (ANN) has been employed in many fields. It has been implemented for Integrated Fire Evolution Monitoring System (IFEMS) [1],for MR brain tumor image classification [2], for handwritten signature verifications [3]; for watershed hydrological modeling [4], and forcustomer relationship management [5]. For more details see [6].

The Adaptive Resonance Theory (ART) Artificial Neural Networks cover both supervised and unsupervised training algorithms. However, the fundamental principle for both forms is measuring the choice value for each committed category node. The maximum choice value node is the candidate to represent the input patterns. When the candidate node fails to pass the vigilance parameter we have to put it in shut off mode and determine the new maximum choice value node. We keep doing this until either a committed category node can represent the current input or a new node must be committed. Such process is time consuming.

II. Objective

The objective of this study is to introduce a new approach that merges the choice value and the testing process against the vigilance parameter ρ in a single step. The first candidate committed category node either can represent the current input or a new node must be committed. This is to avoid testing all committed category node. This reduces the training time, ρ is fixed during training, and eliminates the choice value parameterα. Also, design the supervised version for (Adaptive Pointing Theory) APT the APT-TAG and APT-BAG.

III. Adaptive Resonance Theory

The ART ANN covers both unsupervised training algorithms: Fuzzy ART [7]; Flagged and Compact Fuzzy ART [8], and supervised: Fuzzy ARTMAP [9]; ART-TAG [10]; ART-BAG [11]. However, the fundamental principles for ART ANNs are:

- Measuring the choice value for each committed category node. The choice value represents the activation level for each committed category node;

$$T_j^{(t)} = \frac{\sum_{i=1}^{2M} (A_i^{(t)} \wedge w_{ij})}{\alpha + \sum_{i=1}^{2M} w_{ij}} \quad j = 1, \ldots, C \tag{1}$$

where w_{ij} are the weights between each committed category node j and the input nodes. $A^{(t)} \in [0, 1]$ is the normalized input pattern and its complement, α is the choice value and C is the number of committed category nodes.

- Determine the maximum choice value node J as a candidate to represent the input patterns;

$$T_J^{(t)} = \max\{T_j^{(t)}\},\ j = 1, \ldots, C \tag{2}$$

- Compute the match value S_J for this promising node;

$$S_J = \frac{\sum_{i=1}^{2M} (A_i^{(t)} \wedge w_{iJ})}{M} \tag{3}$$

- Test this promising node against the vigilance parameter ρ[0,1].

If $S_J \geq \rho$ weights for node J are trained:

$$w_{iJ}^{new} = \beta(A_i^{(t)} \wedge w_{iJ}^{old}) + (1-\beta) w_{iJ}^{old}; i = 1, \ldots, 2M \tag{4}$$

where $\beta(0,1]$ is the learning parameter. Otherwise, we have to put node J in shut off mode and determine the new maximum choice value node according to its choice value. M is number of input features.

We keep doing this until either a committed category node can represent the current input or we run out of committed category node (all committed category nodes are in shut off mode) and a new category node must be committed. This is

Manuscript received January 5, 2014; revised March 12, 2014.

Kamal R. Al-Rawi is with Department of Computer Science, Faculty of Information Technology, University of Petra, Jordan (e-mail: kamalr@uop.edu.jo).

Consuelo Gonzalo is with Universidad Politécnica de Madrid, Center for Biomedical Technology (CTB), Pozuelo de Alarcón, Spain.

DOI: 10.7763/IJCCE.2014.V3.322

time consuming during training.

We followed closely the Compact Fuzzy ART. It differ from Fuzzy ART by: 1) No initialization values for the category nodes, 2) No initialization values for the weights, and 3) Only committed category nodes rather than the whole number of category node are involved in determining the winning category node. For more details about Compact Fuzzy ART see [8].

IV. Adaptive Pointing Theory (APT)

We will describe the training algorithm for (Adaptive Pointing Theory) APT in unsupervised form. The architecture for it is the same as that for Compact Fuzzy ART. See Fig. 1 for the full architecture. However, the training algorithm for APT is very efficient relative to Compact Fuzzy ART from training time point of view.

The main steps for training APT ANN is:

- Measuring the match value for each committed category node. The match value S for a committed category node represents the similarity between the input and the committed category node;

$$S_j^{(t)} = \frac{\sum_{i=1}^{2M} (A_i^{(t)} \wedge w_{ij})}{\sum_{i=1}^{2M} (A_i^{(t)} \vee w_{ij})}; \; j = 1;...;C \tag{5}$$

where, $(A \wedge w) = \min(A, w); \; (A \vee w) = \max(A, w)$

The nominator in the above equation can be considered as the intersection of A and W, and the denominator as the union of A and W. When $A = W$, we have value of one which is the max value for the vigilance parameter. When $\rho = 1$, we have a perfect match between the input and the committed category node. The choice value parameter α is not required. It had been inserted to break the tie in choice value between two or more committed category nodes.

- Determine the maximum match value node J as a unique candidate to represent the input patterns;

$$S_J^{(t)} = \max\{S_j^{(t)}\}; \; j = 1,...,C \tag{6}$$

- Test the matching value for the max node against the vigilance parameter ρ. IF $S_J^{(t)} \geq \rho$ the weights for node J is trained using equation-4.

Otherwise, new category node must be committed.

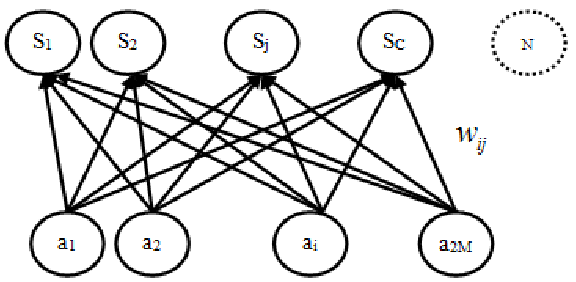

Fig. 1. The architecture of adaptive pointing theory (APT). C is number of committed category nodes. N is the full capacity of the Neural Network. Only committed category nodes are involved in match value calculations. C<<N.

V. Algorithm of APT

A. Training Algorithm of APT

1) Input parameters
   - Dynamic parameters;
     - $\rho \in [0, 1]$: vigilance parameter. ρ=1 for perfect matching.
     - $\beta \in (0, 1]$: The dynamic learning parameter; β=1 for fast learning.
   - Data characteristics;
     - M: The dimension of the input features.
     - Pt: The number of patterns to be used in learning.
   - Initialization;
     - Number of iterations t=1.
     - Number of committed category nodes C=1.
2) New input

$$A_i^{(t)} = \begin{Bmatrix} a_i^{(t)} & for \;\; 1 \leq i \leq M \\ 1 - a_i^{(t)} & for \;\; M + 1 \leq i \leq 2M \end{Bmatrix}$$

3) Compute the match value for each committed category nodes

$$S_j^{(t)} = \frac{\sum_{i=1}^{2M} (A_i^{(t)} \wedge w_{ij})}{\sum_{i=1}^{2M} (A_i^{(t)} \vee w_{ij})}; \; j + 1,...,C$$

4) Determine the Node J, which has the maximum match value

$$S_J^{(t)} = \max\{S_j^{(t)}\}; \; j = 1,...,C$$

5) If $S_J^{(t)} \geq \rho$ train the committed category node J

$$w_{iJ}^{new} = \beta(A_i^{(t)} \wedge w_{iJ}^{old}) + (1 - \beta) w_{iJ}^{old}; i = 1,...,2M$$

Else, Increase committed category node by one;

$$C = C + 1$$

Assign weights of the new node C;

$$w_{iC}^{first} = \beta A_i^{(t)} + (1 - \beta); i = 1,...,2M$$

6) If you have more training pattern GOTO STEP (2)
7) Training has been done. Save the committed category nodes and their weights.

B. Classification Algorithm of APT

1) New input

$$A_i^{(t)} = \begin{Bmatrix} a_i^{(t)} & for \;\; 1 \leq i \leq M \\ 1 - a_i^{(t)} & for \;\; M + 1 \leq i \leq 2M \end{Bmatrix}$$

2) Compute the match value for all committed category nodes

$$S_j^{(t)} = \frac{\sum_{i=1}^{2M} (A_i^{(t)} \wedge w_{ij})}{\sum_{i=1}^{2M} (A_i^{(t)} \vee w_{ij})}; \; j = 1,...,C$$

3) Determine the node J, which has the maximum match value;

$$S_J^{(t)} = \max\{S_j^{(t)}\}; \; j - 1,...,C;$$

4) Node J represents this input.
5) GOTO STEP (1) for next input until all of them are classified.
6) Classification has been done.

VI. Application

The supervision of APT ANN using TAGging (AL-Rawi 1999) and BAGging (AL-Rawi *et al*. 1999) approaches will leads to APT-TAG and APT-BAG, respectively.

In APT-TAG, the weights of the winning committed category node are trained when it passes the vigilance parameter ρ and its TAG equal to the class of the current input as well. Otherwise a new node must be committed and tagged with the class of the current input. Details about TAGging approach can be seen in [10].

In APT-BAG, the weights of the winning committed category node are trained when it passes the vigilance parameter ρ and the BAG of the winning category node is equal to the class of the current input. Otherwise a new node must be committed from the BAG that represents the class of the current input. Details about BAGging approach can be seen in [11].

ART-BAG reduces sharply the training time relative to ART-TAG especially when we have large number of classes and large number of committed nodes [12]. However, training time reduction for APT-TAG is slightly better than APT-BAG since it point to the winning node in a single test. That is because in TAGging approach we find the max for all committed category nodes while for BAGging approach we have to find the max for each BAG then we determine the max among them.

We run APT-TAG to analysis Landsat TM image. The image we used with six bands and 13 different classes. The results are compared to ART-TAG. Both run on an Intel P4-1.7GHz with 256MB RAM. 9,000 exemplars used for training. The total number of pixels is 62,000. Combination of vigilance parameter ρ=0.8, 0.9, 0.98 and dynamic learning rate β=0.3, 0.6, 0.9 are used. β =1.0 is also used for ρ=0.98. The classification accuracy is about the same while number of committed node C for APT-TAG is less than that for ART-TAG. This also reduces both training and classification time. These runs are shown in Table I.

TABLE I: Classification of Landsat Thematic Mapper Image Using APT-TAG and ART-TAG ANNs. The Number of Committed Category Nodes C and the Classification Accuracy are Listed for Different Training Parameters Using 9,000 Exemplars for Training and 62,000 for Testing.

		APT-TAG		ART-TAG	
ρ	β	C	Accuracy%	C	Accuracy%
0.80	0.3	17	73	51	71
	0.6	20	65	62	70
	0.9	24	61	71	70
0.90	0.3	50	78	70	78
	0.6	73	77	80	77
	0.9	100	67	119	77
0.98	0.3	380	81	390	81
	0.6	803	86	807	85
	0.9	1531	86	1511	87
0.98	1.0	2117	88	2172	89

VII. Discussion and Conclusions

The classification accuracy for both APT-TAG and ART-TAG are about the same order. However, the training time for APT-TAG is much less than ART-TAG because a single test is requires determining if the committed category node, with max matching value, can represent the current input or a new node must be committed rather than checking every single committed category node. In the ART-TAG case C checking time is required which is C times more than that for APT-TAG before deciding for a new category node to be committed.

Moreover, the vigilance parameter ρ is fixed during training. The choice value parameter α is eliminated. Number of committed category node is reduced.

Such algorithm will reduce the training time and testing time as well, especially when we have a large number of committed category nodes, which is the case in most practical problem. More studies are required to insure the reduction in number of committed category nodes and classification performance as well.

References

[1] K. R. Al-Rawi, J. L. Casanova, A. Ramo, and E. M. Louakfaoui, “Integrated fire evolution monitoring system (IFEMS) for monitoring spatial-temporal behavior of multiple fire phenomena,” *International Journal of Remote Sensing*, vol. 23, no. 10, pp. 1967-1986, September 2002.

[2] D. J. Hemanth, D. Selvathi, and J. Anitha, “Application of adaptive resonance theory neural network for MR brain tumor image classification,” *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, vol. 5, issue 1, pp. 61-75, January 2010.

[3] T. Dash, S. Chattopadhyay, and T. Nayak, “Handwritten signature verifications using adaptive resonance theory type-2 (ART-2) net,” *Journal of Global Research in Computer Science*, vol. 3, no. 8, pp. 21-25, 2012.

[4] B. Chen, P. Li, and T. Husain, “Development of an integrated adaptive resonance theory mapping classification system for supporting watershed hydrological modeling,” *Journal of Hydrology Engineering*, vol. 17, no. 6, pp. 679–693, June 2012.

[5] M. Anand, Z. Khan, and R. S. Shukla, “Customer relationship management using adaptive resonance theory,” *International Journal of Computer Applications*, vol. 76, no. 6, pp. 43-47, 2013.

[6] S. Grossberg, “Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world,” *Journal of Neural Networks*, vol. 37, pp. 1-47, 2013.

[7] G. A. Carpenter, S. Grossberg, and D. B. Rosen, “Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system,” *Journal of Neural Networks*, vol. 4, issue 6, pp. 759-771, 1991.

[8] K. R. Al-Rawi and C. Gonzalo, “Flagged and compact fuzzy ART: fuzzy ART in more efficient forms,” *International Journal of ACM Jordan, Chapter-ISWSA Research Bulletin*, vol. 1, no. 3, pp. 98-117, 2010.

[9] G. A. Carpenter, S. Grossberg, N. Markuzan, J. H. Reynold, and D. B. Rosen, “Fuzzy artmap: a neural network architecture for incremental supervised learning of analog multidimensional maps,” *IEEE Transaction on Neural Network*, vol. 3, issue 5, pp. 698-882, September 1992.

[10] K. R. Al-Rawi, “Supervised ART-I: A new neural network architecture for learning and classifying multivalued input patterns,” *Lecture Notes in Computer Science*, no. 1606, pp. 756-765, 1999.

[11] K. R. Al-Rawi, C. Gonzalo, and A. Arquero, “Supervised ART-II: A new neural network architecture, with quicker learning algorithm, for classifying multi-valued input patterns,” in *Proc. The European Symposium on Artificial Neural Network ESANN 99*, Bruges, Belgium, 1999, pp. 289-294.

[12] K. R. Al-Rawi, C. Gonzalo, and E. Martinez, ”Supervised ART-II for classification Landsat TM image,” in *Proc. 19th Symposium of European Association of Remote Sensing Laboratories* (*EARSeL*), Valladolid, Spain, pp. 229 –236, 2000.

Kamal R. Al-Rawi was born in Ana, Anbar, Iraq on July, 1, 1953. He holds BS in physics, University of Mosul, Mosul, Iraq in 1977. M.Sc. In meteorology, Oklahoma University, Norman, OK, USA, 1985. Ph.D. in space physics, University of Valladolid, Valladolid, Spain 2000. Ph.D. in computer science, University Polytechnique of Madrid, Madrid, Spain 2001.

He joint Department of Physics, Faculty of Education in Salah Aldeen University, Arbil, Iraq, in 1985-1988. He was with Department of Physics, University of Anbar, Ramadi, Anbar, Iraq, 1988-1996. He was the director of the cultural relation office 1995-1996. He was a Ph.D. student with a scholarship from the Spanish ministry of foreign affairs, Spanish Institute of Cooperation with Arabic World Spain 1996-2001. He taught at Department of Computer Science, Yarmook University, Irbid, Jordan during Spring semester in 2001. He was appointed as assistant professor, Department of Computer Science, Faculty of Information Technology, Al-Ahlia Amman University, Amman, Jordan in 2002-2003. He was an assistant professor of Department of Computer Science, Faculty of Information Technology, University of Petra, Amman, Jordan in 2003-2008. He is an associate professor at University of Petra since 2008 till now. His research interest in theory of artificial neural networks and their applications for analysis satellite images. He has joint the FUEGO Project for the European Space Agency ESA.

Dr. AL-Rawi is a member of ACM. He was a candidate for the best dissertation in the Faculty of Informatics, University Polytechnique of Madrid, Madrid, Spain. He was selected for the Ramon Y Cajal program in the field of computer computation for five years.

Consuelo Gonzalo-Martín received the B.A degree from the Salamanca University and the Ph.D from the Complutense University of Madrid, both in physics, in 1986 and 1989 respectively. Since 1993 she is an assistant professor at the Department of Arquitectura y Tecnología de Sistemas Informáticos in the Facultad de Informática (Universidad Politécnica de Madrid). Her main research areas are image processing and artificial neural networks for application to remote sensing. In particular, she has worked in the development of different algorithms for optical image fusion and ART and SOM artificial neural networks. In September 2012, she joined the Center for Biomedical Technology of the UPM, where she is involved in research and development projects for text and image mining in the health care domain. As a result of her research she has directed more than 10 financed research projects and participated in around 15 more; coauthor of 25 international journal publications, most of them with high impact factor in the Remote Sensing area as the International Journal of Remote Sensing or the Canadian Journal of Remote Sensing between others and has participated 40 international congresses.