01 Jan 1992

SeMi-Supervised Adaptive Resonance Theory (SMART2)

Christopher J. Merz
Missouri University of Science and Technology, merzc@mst.edu

William E. Bond
Missouri University of Science and Technology, bondw@mst.edu

Daniel C. St. Clair

Recommended Citation

C. J. Merz et al., "SeMi-Supervised Adaptive Resonance Theory (SMART2)," *Proceedings of the International Joint Conference on Neural Networks, 1992*, Institute of Electrical and Electronics Engineers (IEEE), Jan 1992.
The definitive version is available at https://doi.org/10.1109/IJCNN.1992.227046

SeMi-supervised Adaptive Resonance Theory (SMART2)[1]

Christopher J. Merz
University of Missouri-Rolla
Rolla, Missouri 65401

Daniel C. St. Clair[2]
University of Missouri-Rolla
Engineering Education Center
St. Louis, Missouri 63121

William E. Bond
McDonnell Douglas
Research Laboratories
St. Louis, Missouri 63166

Abstract

Adaptive Resonance Theory (ART) algorithms represent a class of neural network architectures which self-organize stable recognition categories in response to arbitrary sequences of input patterns. This paper discusses incorporation of supervision into one of these architectures, ART2. Results of numerical experiments indicate this new SeMi-supervised version of ART2 (SMART2) outperforms ART2 for classification problems. The test accuracy of SMART2 is similar to that of Backpropagation. However, SMART2 network structures are easier to interpret than the corresponding structures produced by Backpropagation.

1.0 Introduction

McDonnell Douglas Research Laboratories is interested in using machine learning algorithms to develop classifier systems for use in diagnostic and recognition tasks. In these domains, continuous-valued inputs are used to predict a small set of classes. Backpropagation [Rumelhart et al, 1986] and ART2 [Carpenter and Grossberg 1987, 1988] neural networks have been investigated for these applications. Results have shown that, given a representative training set, Backpropagation can produce a good classification system. Unfortunately, network configurations are not obvious, and large networks usually require considerable time to train.

One type of Adaptive Resonance Theory (ART) algorithm, ART2, performs unsupervised learning by attempting to discover patterns or categories in input data by grouping similar input patterns. Unfortunately, similar training patterns may be grouped together regardless of the class they represent. This often results in categories describing multiple classes.

A modified version of ART2 was developed which uses class information to supervise the construction of ART2 categories. This supervised version of ART2, SeMi-supervised Adaptive Resonance Theory (SMART2), uses an input pattern's class to allow self-organization only among patterns of the same class. Thus, similar input patterns from different classes are not allowed to interfere in the development of categories. Training was also improved by increasing the sensitivity of the network only for patterns which the network has difficulty classifying correctly. The resulting algorithm has classification accuracy similar to Backpropagation but requires significantly less training time. In addition, SMART2 categories are easy to interpret.

The next section outlines the ART2 class of networks and describes two new design principles which can be used to supervise the learning process. Section three describes a new paradigm, SMART2. The results and analysis of runs on several data sets by SMART2, ART2, and Backpropagation are analyzed in the fourth section.

2.0 ART2

An ART2 architecture is composed of two major parts; the attentional subsystem and the orienting subsystem. The objective is to sort input patterns into categories of "similar" patterns. Figure 1 depicts a SMART2 architecture. (The basic ART2 architecture is obtained by removing the Test History and Class entities from this architecture.)

The orienting subsystem is an optional part of the network which, when active, helps guide the attentional subsystem in its search for the category bearing the most resemblance to the current input pattern. When searching the existing category structures of an ART2 network, the vigilance level, ρ, designates the minimum strength of a match. Training begins when a category is found whose match with the input pattern is a value greater than ρ. The order in which candidate categories are searched is controlled by the attentional subsystem. Categories which fail to

---

[1] Work supported by the McDonnell Douglas Independent Research and Development program.
Export Authority: 22 CFR 125, 4(b) (13).

[2] D. C. St. Clair is a Visiting Principal Scientist at McDonnell Douglas Research Laboratories in St. Louis.

0-7803-0559-0/92 $3.00 © 1992 IEEE

provide an adequate match are not considered as candidates for refinement. A new category is formed when there are no more established category candidates. Higher values of ρ produce finer categories. When $\rho = 0$, the orienting subsystem is considered inactive and the network is allowed to completely self-organize.

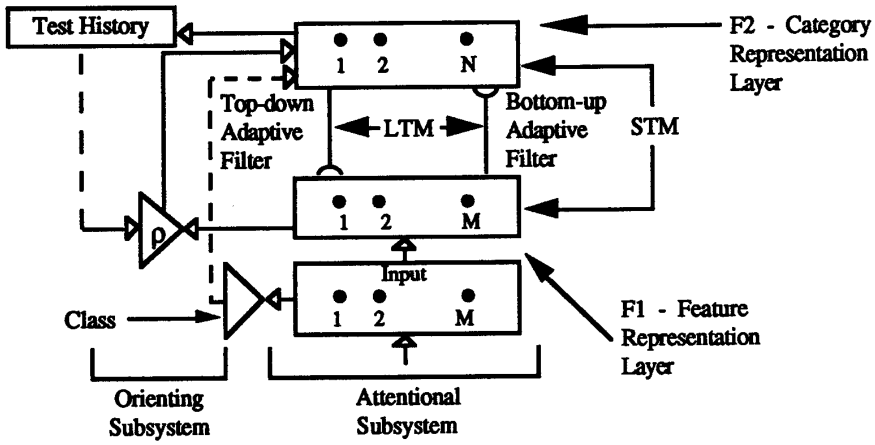

Figure 1. Basic SMART2 Architecture

The attentional subsystem is composed of two parts; short term memory (STM) and long term memory (LTM). The two STM subcomponents are F1, the feature representation layer, and F2, the category representation layer. The F1 layer produces contrast enhancement, noise suppression, normalization, and pattern matching. It is composed of three layers of M nodes, where M is the number of inputs in the training patterns. The top layer of F1 feeds the bottom-up adaptive filter of LTM. In addition, it reads and normalizes top-down filtered input from LTM. The bottom layer of F1 reads and normalizes input patterns and the middle layer matches patterns from top and bottom layers before sending a composite pattern through to the other two layers. Each node in F2 denotes a category. This layer evaluates competitive interactions of the nodes/categories and chooses the most active F2 node in response to an F1 pattern being applied to the bottom-up adaptive filter. The F2 layer also suppresses reset nodes, as directed by the orienting subsystem. Reset nodes are nodes which provide poor matches based on the current ρ setting.

LTM is made up of two components, the bottom-up adaptive filter and the top-down adaptive filter. Their collective purpose is to store the category structures. In the bottom-up adaptive filter, each of the M nodes in the top layer of F1 are connected, by weighted arcs, to each of the N nodes in F2. When a pattern is first applied to the network, it goes through the F1 layer and primes the F2 layer via the bottom-up adaptive filter. At this point, the F2 category whose bottom-up adaptive filter has the highest dot product with the input pattern is chosen as the next candidate for a match. This category's image or top-down expectation is projected down to the top F1 layer via the top-down adaptive filter. Then the top and middle layers of the F1 layer are checked for an adequate match by the orienting subsystem. If the candidate F2 category does not provide a satisfactory match, that node is reset and the search continues. Resetting a node prevents it from competing again. Once an adequate match is found in the F2 layer, its image is projected back down to the F1 layer. The F1 layer uses outputs from the top-down adaptive filter in LTM and outputs from the Input layer to determine the next modification of LTM.

Since ART2 is unsupervised, it makes no use of the class information associated with a training pattern. This can cause two or more training patterns with similar input vectors but different classes to activate the same category. If a training pattern of a certain class were not allowed to influence categories established for other classes, the network would be less likely to develop "compromise" categories which attract patterns of more than one class. Thus, the first new design principle to be added to ART2 was:
DP1: Allow only categories of the same class to compete for an input training pattern.

In addition, the vigilance level for ART2 remains constant during training. High vigilance levels cause unnecessary expansion of categories since they cause ART2 to detect insignificant fluctuations among the patterns. If a high vigilance setting is needed to distinguish between patterns of certain classes, that vigilance level is imposed on all of the classes. To overcome this, a second design principle was added: ***DP2: Use high vigilance only for***

patterns which are difficult to classify. Carpenter and Grossberg's ARTMAP [1991] uses a more complex technique to adjust vigilance.

3.0 SMART2

The integration of these new design principles into the existing ART2 architecture results in a more sophisticated orienting subsystem. Under the new algorithm, SMART2, self-organization is allowed only among patterns of the same class. In addition, patterns previously misclassified require a strong match before training can begin. Figure 1 shows the basic SMART2 structure. The Class and Test History entities correspond to DP1 and DP2 respectively. Figure 2 outlines the SMART2 algorithm. Sections shown in italics are additions to the original ART2 algorithm [Merz 1990]. The only change to the attentional subsystem was in the category representation layer, F2. During training, SMART2 uses class information from the current training pattern to inhibit nodes associated with other classes. Thus, as each new category is formed, it is tagged with the class of the training pattern that formed it.

Two new functions were also added to the orienting subsystem. The first function resets all F2 nodes with class tags other than the class of the current training pattern (Step 2.1). This enforces DP1. The second function tracks misclassified training patterns during the test cycle, and then increases ρ for those patterns during the next training phase (Step 0). This last change implements DP2 by requiring finer categories through higher vigilance for classes with patterns which are difficult to classify.

The algorithm was also modified in order to control the amount of time a pattern remained on the network. Originally, a pattern was removed after a specified number of modifications to LTM. The SMART2 algorithm removes a pattern whenever the last LTM adjustment does not significantly affect the strength of the match. For a weight adjustment to be considered significant, experiments indicate the match strength must change by at least .0000001 per training iteration. This modification reduces training time and prevents overfitting of the categories.

Start with low vigilance (usually $\rho = 0$)

FOR (each training pattern)
  WHILE (a training pattern is applied to network) DO

0. *IF (the network misclassified this pattern on the last testing cycle)*
  THEN turn vigilance up (usually to $\rho = .999$),
1. F1 processing,
2. Apply F1 results to Bottom-up Adaptive Filter,
2.1. *Reset all F2 nodes which are associated with a class that differs*
  from the class of this training pattern,
3. F2 processing (choose winning category, j, or add a new category),
4. Send down top-down expectation from category j via the Top-down Adaptive Filter.
5. IF (match strength between bottom-up inputs and top-down expectation > ρ)
  THEN
    Adjust the Long Term Memory (LTM) to make the Bottom-up and
    Top-down Adaptive Filters look more like the input pattern,

    IF (this category has not been used before)
    THEN tag it with this training pattern's class,

    IF *(weight adjustment does not provide a 'significant' improvement in the match)*
      OR (maximum number of weight adjustment is exceeded)
    THEN remove the pattern from the network,
  ELSE
    Activate the Orienting Subsystem to reset the jth category in F2 (i.e., inhibit it from competing again while this pattern remains on the network).

FOR (each training pattern)
  Apply training pattern to the new version of the network allowing any category to win. See if the class of the pattern is the same as the tag of the category it mapped to.

Figure 2. The SMART2 Algorithm (One training iteration)

4.0 Numerical Evaluation of SMART2

Several experiments were conducted to evaluate SMART2 against two different types of neural network algorithms: ART2 and Backpropagation (BP). In order to examine the effect of each of the SMART2 design principles, three versions of SMART2 were tested; SMART2-I which consisted of ART2 + DP1, SMART-II which consisted of ART2 + DP2, and SMART2 which is ART2 + both new design principles. The notation BP (x-y-z) denotes a BP network with x input nodes, y nodes in the hidden layer, and z nodes in the output layer.

Four difficult domains were chosen on which to test the algorithms. The first three domains represent real data while the fourth was simulated data. These datasets are characterized in Table I:

1) **Nondestructive evaluation (NDE7)** is performed on structures such as those found in aircraft in order to identify subsurface defects in metallic and composite parts. The NDE7 dataset consisted of 360 instances; each instance has 94 continuous-valued attributes corresponding to the strength of the ultrasonic signal plus an associated class value. There are seven possible class values; six different defects and no defect [Bond et al 1992].
2) **Glass** left at the scene of a crime often can be used as evidence if it can be correctly identified. This dataset, from B. German, came from the University of California-Irvine's Repository of Machine Learning Data. The objective is to classify six glass types by using nine attributes representing chemical and physical properties.
3) **Radar range data** was taken from actual radar outputs. This type of ship classification problem is described in Breiman et al [1984].
4) **Waveform recognition** data is simulated data representing the problem of recognizing three waveforms. This problem is difficult because random Gaussian noise was added to the waveforms [Breiman et al 1984].

Dataset	Number Instances	No. Input Attributes	Number Classes	Dataset	Number Instances	No. Input Attributes	Number Classes
NDE7	360	94	7	Radar	1200	64	6
Glass	214	9	6	Waveform	400	21	3

Table I. Dataset Summary

Experiments were performed on each dataset by randomly selecting 70% of the data for use as a training set. The remaining 30% was used for testing. This procedure was repeated five times for each problem domain. Table II values represent the average of five experiments and shows the performance of the algorithms on the various datasets. Backpropagation architectures were selected to optimize BP performance. All algorithms were trained exhaustively. Both the highest training accuracy and the Epoch where this highest training accuracy occurred were recorded. The number of training Epochs required for the ART2/SMART2 family is 2-3 orders of magnitude less than BP in several cases!

The Number of Categories and Number of Multiple Class Categories columns contain information about the categories identified by each algorithm. For example, for the Radar dataset, SMART2 and SMART2-I created an average of 67.6 and 165.2 categories respectively. Since both SMART2 and SMART2-I utilize DP1, only categories which map to a single class can be created. The radar data contained 6 different classes. SMART2 and SMART2-I produced multiple categories for each class. Thus, each category can be viewed as a subclass. Examination of the training instances associated with each of the categories verify that this is true! Data plots for instances in the same category appear to be more similar than data plots from other categories for the same class.

In addition, plotting training instances associated with each of the ART2 categories indicates that the high vigilance required to separate some similar ship patterns also causes unnecessary category expansion. This results in the formation of extra categories which may not correspond to specific classes. The Number of Categories and the Number of Multiple Class Categories in Table II demonstrate this. SMART2 and SMART2-II avoided this unnecessary expansion of the categories by using the dynamic vigilance design principle DP2.

Both ART2 and SMART2-II form categories based on the ART2 similarity measure without regard for class value. For example, the average number of Radar categories found by ART2 was 164.4. Of these, an average of 12 categories could be associated with more than one class value. This class independent formation of categories affects accuracy since accuracy measurements are based on the class associated with a category. ART2 and SMART2-II may actually find relationships in data which are different from those found by SMART2 and SMART2-I. This makes ART2 and SMART2-II useful in data exploration.

Dataset	Algorithm	Epoch where maximum accuracy occurs	Training Accuracy	Test Accuracy	Number Categories	Number Multiple Class Categories
NDE7	ART2	2.0	100	81.9	96.8	0
	SMART2	3.0	100	98.2	17.4	0
	SMART2-I	1.8	100	98.7	109.2	0
	SMART2-II	5.0	98.5	98.4	45.4	0.4
	BP (94-10)	3,759.4	100	99.1	10.0	0
Glass	ART2	9.8	61.1	42.2	47.6	10.6
	SMART2	13.8	77.3	50.3	43.4	0
	SMART2-I	6.8	84.3	55.6	57.0	0
	SMART2-II	11.6	58.8	42.5	35.4	9.8
	BP (9-10-7)	6,711.4	86.7	62.2	7.0	0
Radar	ART2	9.6	94.8	94.2	164.4	12
	SMART2	12.6	99.3	96.7	67.6	0
	SMART2-I	11.0	98.4	96.0	165.2	0
	SMART2-II	13.6	93.9	92.8	68.4	13
	BP (64-12-6)	1,441.2	99.6	98.1	6.0	0
Waveform	ART2	1.0	94.9	85.8	200.0	11.6
	SMART2	12.8	100.0	86.8	86.0	0
	SMART2-I	3.4	99.4	88.3	174.4	0
	SMART2-II	14.8	99.8	86.8	133.4	0.4
	BP (21-10-3)	910.4	99.9	90.1	3.0	0

Table II. Performance of Algorithms on Different Datasets

From the results shown in Table II, adding DP1 to ART2 appears to improve the test accuracy of the network and, in some cases, increases the number of categories. Adding DP2 to ART2 slightly improved the test accuracy in three of the four datasets tested and dramatically reduced the number of categories in all data sets. Adding both design principles to ART2 has the net effect of increasing test accuracy and reducing the number of categories. In addition, if the number of categories created is greater than the number of class values, then subclasses have been identified.

As expected, BP consistently produces the best test accuracy. However, BP networks require extensive training. In addition, the results are difficult for humans to interpret.

All of the experimentation for this paper was done on a busy multi-user system which did not allow for precise time comparisons. However, Table III cites two statistics which help quantify the time complexity of the various ART algorithms. In particular, the column indicating the number of resets is most correlated with the approximate run-times observed when all four algorithms were started simultaneously on the same data set. The rough ordering for the number of resets and the observed run-times is **Smart2 < Smart2-I << Smart2-II < Art2 << BP** where the placement of BP is based solely on the approximate observed run times.

Dataset	# Resets	# Weight Adjust.	Dataset	# Resets	# Weight Adjust.
NDE7			**Radar**		
ART2	1,512.2	31,533.4	ART2	438,368.6	37,988.8
SMART2	16.6	50.699.6	SMART2	4,763.0	47,040.2
SMART2-I	1,279.6	31,086.2	SMART2-I	68,267.6	37,562.2
SMART2-II	12,442.6	46,108.8	SMART2-II	77,394.8	46,770.4
Glass			**Waveform**		
ART2	2,636.8	8343.8	ART2	329,596.6	1,400.0
SMART2	513.2	9040.2	SMART2	1,799.4	65,637.4
SMART2-I	784.2	8637.8	SMART2-I	6,357.8	23,849.6
SMART2-II	1,799.2	8779.6	SMART2-II	7,853.0	5,324.1

Table III. Work Analysis

Table III shows that SMART2 required considerably more weight adjustments on the NDE7 dataset than any other algorithm. However, the ART2 and SMART2-II run times were approximately 6 times longer or more due to the relatively high number of resets. One weight adjustment consists of some F1 processing and the multiplication of a single bottom-up vector and a single top-down vector. A reset operation requires the same amount of F1 processing plus the multiplication of all vectors in the bottom-up adaptive filter and the projection and evaluation of the winning top-down pattern.

Two major drawbacks of ART architectures are their poor ability to deal with noise and the complexity of their implementation and tuning [Caudill and Butler 1990]. The three data sets (NDE7, Radar, and Waveform) which have a significant amount of noise illustrate that the new design principles work together to overcome this problem by forming simpler categories with greater accuracy and less training. The two new design principles increase implementation complexity. However, the need to tune the parameters of the network is reduced.

5.0 Conclusions

The SMART2 algorithm was created by adding two new design principles to ART2. The first design principle, DP1, insures that only categories of the same class compete for an input training pattern. The second design principle, DP2, reduces the number of categories formed by controlling vigilance. Experimental results indicate that SMART2 is more computationally efficient, more accurate, and more tolerant of noisy inputs than ART2.

Although test accuracy is slightly better for Backpropagation networks than it is for SMART2/ART2 networks, SMART2/ART2 networks are much easier to interpret. Training times for SMART2/ART2 algorithms appear to be considerably less than those for Backpropagation. In addition, normalization of input data is not required for SMART2/ART2 algorithms.

In situations where data exploration is needed and class values are not available, unsupervised learning such as that performed by ART2 and SMART2-II should be used. However, when supervised learning is required, the understandability of SMART2 networks coupled with the small training times required may make SMART2 more useful than Backpropagation.

Acknowledgement

Our special thanks to Steve Aylward, McDonnell Aircraft Company Neural Network Laboratory, for performing the BP calculations.

References

Bond, W.E., St. Clair, D.C., Amirfathi, M.M., Merz, C.J., and Aylward, S., *Neural Network Analysis of Nondestructive Evaluation Patterns*, **1992 Symposium on Applied Computing**, Kansas City, MO, March 1992.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., **Classification and Regression Trees**, Wadsworth & Brooks/Cole, 1984.

Carpenter, G. A., and Grossberg, S., *ART 2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns*, **Applied Optics**, Dec. 1, 1987, No. 23, pp. 4919-4930.

Carpenter, G. A., and Grossberg, S., *The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network*, **Computer: Special issue on Artificial Neural Systems**, Vol. 21, 1988, pp. 77-88.

Carpenter, G. A., Grossberg, S., and Reynolds, J. H., *ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network*, **Neural Networks**, Vol. 4, 1991.

Caudill, M., and Butler, C., **Naturally Intelligent Systems**, The MIT Press, 1990.

Merz, C. J. *ART 2 Architectures: A Survey*, **University of Missouri-Rolla Intelligent Systems Center Technical Report ISC-TR-90-012**, May 1, 1990.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., *Learning Interior Representation by Error Propagation*, **Parallel Distributed Processing**, Vol. 1, Ch. 8, D.E. Rumelhart and J.L. McClelland eds., MIT Press Cambridge, Mass., 1986.