Fuzzy ARTMAP, Slow Learning, and Probability Estimation

Gail A. Carpenter[1], Stephen Grossberg[2], and John H. Reynolds[3]
Department of Cognitive and Neural Systems and Center for Adaptive Systems
111 Cummington Street, Boston University, Boston, MA 02215, USA

Abstract

A nonparametric probability estimation procedure using the fuzzy ARTMAP neural network is here described. Because the procedure does not make *a priori* assumptions about underlying probability distributions, it yields accurate estimates on a wide variety of prediction tasks. Fuzzy ARTMAP is used to perform probability estimation in two different modes. In a 'slow-learning' mode, input-output associations change slowly, with the strength of each association computing a conditional probability estimate. In 'max-nodes' mode, a fixed number of categories are coded during an initial fast learning interval, and weights are then tuned by slow learning. Simulations illustrate system performance on tasks in which various numbers of clusters in the set of input vectors mapped to a given class.

Fuzzy ARTMAP for Probability Estimation

Many pattern recognition applications require an estimate of the probability that an input belongs to a given class. In a medical database, for example, a set of diagnostic measurements are used to estimate the probability that a patient will require a long stay in the hospital. Here, different combinations of variables may be associated with a single output so no single group of variables is predictive. Fuzzy ARTMAP [4, 5] is a neural network that provides a means for automatically selecting complex combinations of factors on which to build accurate probability estimates for application to problems such as medical prediction [1, 7].

We here develop a procedure that uses this architecture for probability estimation. Simulations demonstrate that the method is robust, performing well in applications with widely varying input probability distributions. Unlike parametric probability estimators, fuzzy ARTMAP does not depend on *a priori* assumptions about the underlying data. Two variants of this method are described: the 'slow-learning' mode and the 'max-nodes' mode. A simulation example shows fuzzy ARTMAP's performance in max-nodes mode on a difficult probability estimation

[1]Supported in part by British Petroleum (89-A-1204), DARPA (AFOSR 90-0083 and ONR N00014-92-J-4015), the National Science Foundation (NSF IRI 90-00530), and the Office of Naval Research (ONR N00014-91-J-4100).

[2]Supported in part by the Air Force Office of Scientific Research (AFOSR 90-0175), DARPA (AFOSR 90-0083 and ONR N00014-92-J-4015), and the Office of Naval Research (ONR N00014-91-J-4100).

[3]Supported in part by the Air Force Office of Scientific Research (AFOSR 90-1075), DARPA (ONR N00014-91-J-4100 and ONR N00014-92-J-4015), and the National Science Foundation (NSF IRI-90-00530).

task, in which each of two classes has 97 clusters arranged in two concentric nested spirals.

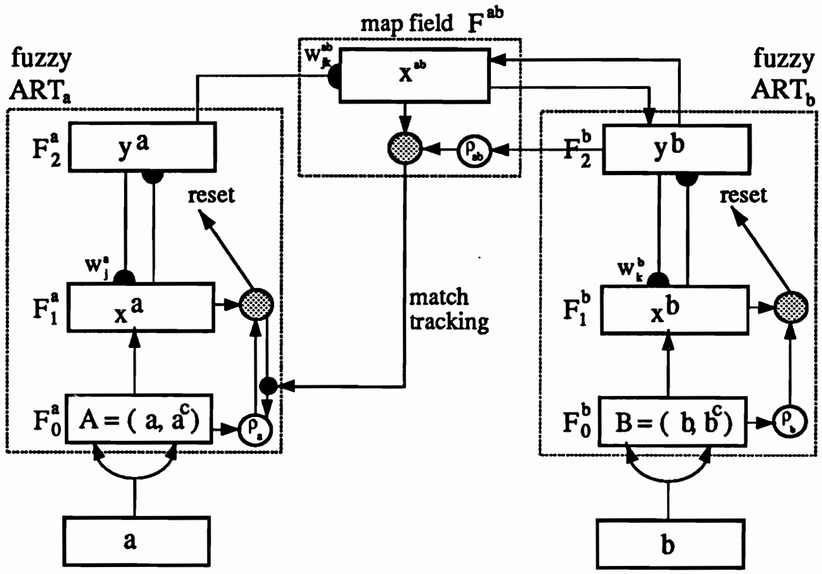

Figure 1. Fuzzy ARTMAP architecture [4]. The ART$_a$ complement coding preprocessor transforms the M_a-vector **a** into the $2M_a$-vector $\mathbf{A} = (\mathbf{a},\mathbf{a}^c)$ at the ART$_a$ field F_0^a. Vector **A** is the input to the ART$_a$ field F_1^a. Similarly, the input to F_1^b is the $2M_b$-vector $\mathbf{B} = (\mathbf{b},\mathbf{b}^c)$. When a prediction by ART$_a$ is disconfirmed at ART$_b$, inhibition of map field activation induces the match tracking process. Match tracking raises the ART$_a$ vigilance (ρ_a) to just above the F_1^a-to-F_0^a match ratio $|\mathbf{x}^a|/|\mathbf{A}|$. This triggers an ART$_a$ search that leads to activation of either an ART$_a$ category that correctly predicts **b** or to a previously uncommitted ART$_a$ category node.

Figure 1 shows the fuzzy ARTMAP architecture [4]. Each fuzzy ARTMAP system includes a pair of Adaptive Resonance Theory modules (ART$_a$ and ART$_b$) that create stable recognition categories in response to arbitrary sequences of input patterns. During supervised learning, ART$_a$ receives a stream $\{\mathbf{a}^{(p)}\}$ of input patterns and ART$_b$ also receives a stream $\{\mathbf{b}^{(p)}\}$ of patterns, where $\mathbf{b}^{(p)}$ is the correct prediction given $\mathbf{a}^{(p)}$. These modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time. The controller is designed to create the minimal number of ART$_a$ recognition categories, or 'hidden units,' needed to meet accuracy criteria.

Parameter ρ_a calibrates the minimum confidence that ART$_a$ must have in a recognition category, or hypothesis, activated by an input $\mathbf{a}^{(p)}$ in order for ART$_a$ to accept that category, rather than search for a better one through an automatically controlled process of hypothesis testing. Lower values of ρ_a enable larger categories to form. These lower ρ_a values lead to broader generalization and a higher degree of code compression. A predictive failure at ART$_b$ increases ρ_a by the minimum amount needed to trigger hypothesis testing at ART$_a$, using a mechanism called *match tracking*. Match tracking sacrifices the minimum amount of generalization necessary to correct a predictive error. Hypothesis testing leads to the selection of a new ART_a category, which focuses attention on a new cluster of $\mathbf{a}^{(p)}$ input features that is better able to predict $\mathbf{b}^{(p)}$. Match tracking allows a single ARTMAP system to learn a different prediction for a rare event than for a cloud of similar frequent events in which it is embedded.

Fuzzy ARTMAP can perform probability estimation in two modes. In slow-learning mode, the system grows incrementally until it achieves a good fit to the underlying probability density function. In max-nodes mode, the user specifies an upper bound on network size. After it has reached this size the network stops growing, but additional training data can still be incorporated into the existing network to improve its probability estimates.

Slow-Learning Mode

In slow-learning mode, fuzzy ARTMAP slowly updates its weights to estimate the probability that an input belongs to a given output class. In particular, when an input activates an ART_a category, the size of the weight from F_2^a to a map field category node (Figure 1) constitutes an estimate of the probability that the input belongs to that category. During supervised learning, the strength of the weight projecting from the selected ART_a category to the selected ART_b category is increased, while the strengths of the weights to other ART_b categories are decreased. A system vigilance parameter, ρ_{ab}, calibrates the degree of surprise, or predictive mismatch, necessary to trigger the search for a different ART_a category. If the weight projecting from the active ART_a category to the active ART_b category is smaller than ρ_{ab}, i.e., if the system is 'surprised' by an unexpected outcome, match tracking triggers a search at ART_a.

Once an ART_a category (J) is chosen whose prediction of the actual ART_b category is strong enough, match tracking is disengaged, and resonance occurs at ART_a. During resonance, learning occurs at ART_a according to the fuzzy ART fast learning equations [6] but slow learning occurs at the map field. Map field learning obeys the equation:

$$(w_{jk}^{ab})^{new} = \begin{cases} (1-\beta_{ab})(w_{jk}^{ab})^{old} + \beta_{ab}x_k^{ab} & \text{if } j = J \\ (w_{jk}^{ab})^{old} & \text{if } j \neq J \end{cases} \tag{1}$$

where x_k^{ab} is the activity of the k^{th} map field node and w_{Jk}^{ab} is the map field weight projecting to map field node k from the active ART_a node J. The map field learning parameter β_{ab} determines the rate of change of the map field weights. Small values of β_{ab} cause the system to base its probability estimate on a long-term average of its experience, while values of β_{ab} near 1 allow adaptation to a rapidly changing environment.

Max-nodes Mode

ARTMAP can also operate in a 'max-nodes' mode, in which the user specifies the maximum number of F_2^a category nodes. Here, map field vigilance ρ_{ab} is set to

1 during early training then lowered to 0 when the maximum number of ART_a categories has been reached. With $\rho_{ab} = 0$, match tracking never occurs in response to a predictive mismatch. With $\rho_{ab} = 1$, match tracking will be triggered whenever a predictive error occurs. The initial 'critical period', when $\rho_{ab} = 1$, establishes a tessellation of the input space associating regions with one output class each. After ρ_{ab} is lowered to 0, weights slowly adjust their estimates of the probability that a member of a given ART_a category belongs to a given ART_b class. With ρ_{ab}=0, the rapid partition established with $\rho_{ab} = 1$ is fine tuned via slow learning.

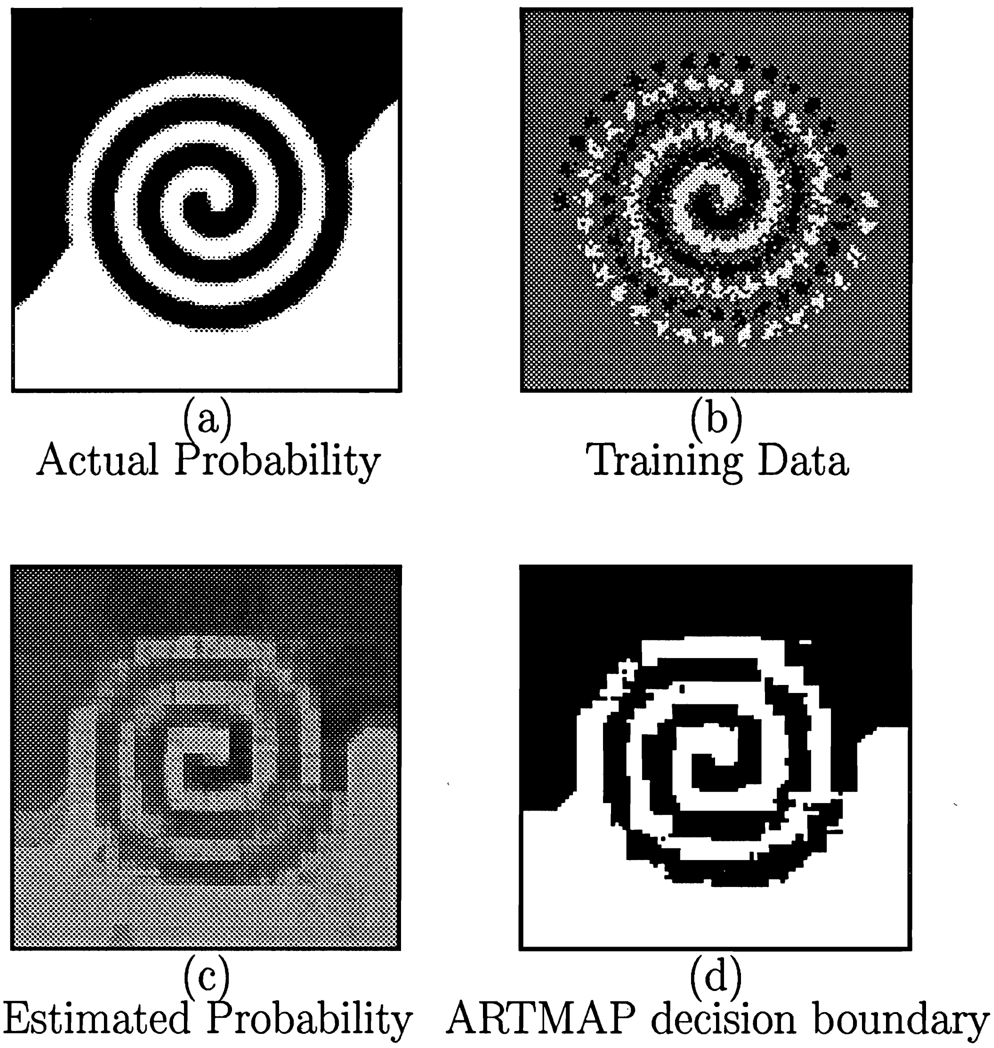

Figure 2. Fuzzy ARTMAP estimated conditional probabilities and decision boundary for a 194-gaussians problem. (a) Actual conditional probabilities. Points falling in lighter areas are more likely to belong to class 1; darker areas, class 2. (b) Actual training data. (c) Fuzzy ARTMAP estimated conditional probability. (d) Fuzzy ARTMAP decision boundary.

Simulation: Noisy nested spirals

Figure 2 shows the performance of the system on a difficult learning problem, in which patterns are drawn from 194 gaussians whose means fall along two nested spirals. Figure 2a shows the actual probability that a pattern falling at each point in the unit square will belong to each of the two classes. Patterns falling in lighter regions are more likely to belong to class 1, while those in darker regions are more likely to belong to class 2. These probabilities represent the best possible estimate calculated using bayes' rule. Figure 2b shows the actual training data, which were drawn from the two distributions with equal probability. White points belong to class 1 and black points belong to class 2. Twenty patterns were drawn from each gaussian, for a total of 1940 patterns belonging to each class. Figure 2c shows the average probability estimate of the ARTMAP model in max-nodes mode, averaged over 10 independent orderings of the training data. On average, the system created 75 ART_a categories. It correctly extracts the shape of the underlying spirals and assigns darker color to the upper left region and lighter color to the lower right. Figure 2d shows decision boundary which results from assigning regions to the class with the higher estimated *a posteriori* probability.

References

[1] Carpenter, G.A., Egbert, D.D., Goodman, P.H., Grossberg, S., Hartz, A.J., Kaburlasos, V.G., Reynolds, J.H., and Rosen, D.B. (1992). *Fuzzy ARTMAP neural network compared to linear discriminant analysis prediction of the length of hospital stay in patients with pneumonia.* Submitted for publication.

[2] Carpenter, G.A. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, **37**, 54–115.

[3] Carpenter, G.A., Grossberg, S. and Iizuka, K. (1992). Comparative performance measures of fuzzy ARTMAP, learned vector quantization, and back propagation for handwritten character recognition. *Proceedings IJCNN-92.* Piscataway, NJ: IEEE. pp. **I**-794-799.

[4] Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, **3**, 698–713.

[5] Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, **4**, 565–588.

[6] Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, **4**, 759–771.

[7] Goodman, P.H., *et al.* (1992) Fuzzy ARTMAP neural network prediction of heart surgery mortality. *Proceedings of the Wang Conference on Neural Networks for Learning, Recognition and Control*, p. 48.