AN EMPIRICAL STUDY OF FUZZY ARTMAP APPLIED TO CYTOGENETICS

Boaz Lerner and Boaz Vigdor

Pattern Analysis and Machine Learning Laboratory
Department of Electrical & Computer Engineering
Ben-Gurion University, Beer-Sheva, Israel
{boaz,vigdorb}@ee.bgu.ac.il

ABSTRACT

The fuzzy ARTMAP (FAM) neural network is evaluated in a pattern classification task of discriminating signals identifying genetic diseases. The FAM provides incremental learning necessary to cope with the expansion of genetic applications and variety of biological preparation techniques. Two training modes of the FAM, training until completion and training with validation, are experimentally compared with respect to their accuracy and sensitivity to the vigilance parameter. Although overfitting the training set, the FAM accuracy on the test set after being trained until completion outperforms that achieved utilizing a validation set. This classification accuracy is completed employing less than five epochs compared to hundreds of training epochs required for other neural network paradigms to accomplish similar performance.

1. INTRODUCTION

Fluorescence in-situ hybridization (FISH) has proven to be a useful tool for the analysis of human chromosomes in clinical diagnosis and genetic research, enabling detection of chromosome numerical abnormalities during normal cell interphase. One of the most common applications of FISH is the enumeration of signals (known as dot counting) within a population of cells, as the signals represent the inspected chromosomes. Dot counting is used for detection and analysis of numerical chromosomal aberrations in prenatal inspection and diagnosis of tumors, as well as demonstration of disease-related chromosomal translocation [1]. One major restriction of the FISH technique is that a large number of cells are needed to be scanned in order to get accurate estimation of chromosome distribution over cell population. Manual evaluation of cells by cytogenetic experts is laborious and time-consuming, therefore it is natural to pursuit the automation of dot counting.

In this work, the fuzzy ARTMAP (FAM) neural network [2] is evaluated in classifying FISH signals as a means to accomplish dot counting. Signals are classified as real or artifact of two genetic diseases, Down syndrome (trisomy 21) and Patau syndrome (trisomy 13). The FAM is a fast incremental learning algorithm that can tackle the dynamics of the cytogenetic environment switching between FISH applications and biological protocols. This benefit of the FAM cannot be found in other neural network or machine learning methodologies previously applied to the FISH domain [3]. The organization of the paper follows. The architecture and dynamics of the fuzzy ARTMAP are summarized in Section 2. The experiments and results are presented in Section 3 and a discussion of these results and future research are given in Section 4.

2. FUZZY ARTMAP ARCHITECTURE AND DYNAMICS

The fuzzy ARTMAP network [2] incorporates two fuzzy ART modules denoted as ART_a and ART_b and linked by a map field module, which associates nodes from ART_a with nodes in ART_b. Each fuzzy adaptive resonance theory (ART) module belongs to a family of networks for fast incremental unsupervised learning. In order to avoid category proliferation all inputs are normalized through a complementary coding scheme. First, the *M*-dimensional input vector $\boldsymbol{I}= (I_1, ..., I_M)$ is normalized so each element I_i belongs to the interval [0,1]. Then, the normalized input is extended to a *2M*-dimensional vector as follows

$$\boldsymbol{I}^e = (\boldsymbol{I}, \boldsymbol{I}_c) = (I_1,...,I_M,1\text{-}I_1,...,1\text{-}I_M). \qquad (1)$$

The fuzzy ART module performs unsupervised learning via clustering of the input samples into categories each forming a hyper-rectangle region in the *M*-dimensional input space. Learning is performed by creation of new categories or by adaptation of existing ones, where category j is completely defined by the set of weights $\boldsymbol{w}_j = (w_{j1}, ..., w_{j2M})$. This categorization is performed in two stages: category choice and vigilance test. In the category choice stage, a choice function is calculated for the current complemented-coded

input and each existing category

$$T_j = \frac{|\boldsymbol{I}^e \wedge \boldsymbol{w}_j|}{\alpha + |\boldsymbol{w}_j|} \tag{2}$$

where $\alpha > 0$ is a choice parameter, $\wedge$ is the fuzzy AND operation and the norm is the L_1 norm

$$|\boldsymbol{x}|_1 = \sum_{i=1}^{M} |x_i| = \sum_{i=1}^{M} x_i. \tag{3}$$

The chosen category is the one achieving the highest value of the choice function (and therefore this stage is often referred to as the competitive stage). When a category J is chosen, a hypothesis test called vigilance test is performed in order to measure the category match to the input $\boldsymbol{I}^e$. If the following match function exceeds the vigilance parameter $\rho \in [0, 1]$

$$\frac{|\boldsymbol{I}^e \wedge \boldsymbol{w}_J|}{|\boldsymbol{I}^e|} \geq \rho \tag{4}$$

the chosen category wins and learning is performed. Else, the current category is removed from the search by forcing T_J to be zero for the rest of this input presentation. As a result, a new category is chosen and the process is continued until an existing category will satisfy the hypothesis test. If none of the existing categories meets the vigilance test, a new category is formed and learning is performed without a vigilance test. Learning is performed by updating the weight vector of the winning category according to

$$\boldsymbol{w}_J^{new} = \beta(\boldsymbol{I}^e \wedge \boldsymbol{w}_J^{old}) + (1 - \beta)\boldsymbol{w}_J^{old} \tag{5}$$

where $\beta \in [0, 1]$ is the learning rate. Fast learning is defined when β is set to one. Learning can be interpreted geometrically as the extension of the category region towards the input sample. The vigilance parameter controls the similarity to the input sample required from the chosen category, thus lowering the vigilance provides broader generalization. In pattern recognition tasks, the input $\boldsymbol{I}_a^e$ to ART_a is the complement coded sample feature vector and the input $\boldsymbol{I}_b^e$ to ART_b is the complement coded sample label (in the test phase no input is inserted to ART_b). As ART_b inputs are labels, ART_b's vigilance parameter ρ_b is configured to one, so each label is clustered by a specific ART_b category. The map field includes a matrix of weights $\boldsymbol{w}^{ab}$ which maps ART_a categories to ART_b categories. The Jth row vector of $\boldsymbol{w}_J^{ab}$ denotes the prediction of ART_b categories as a result of the Jth winning category in ART_a. The map field is activated to produce the output

$$\boldsymbol{x}_{ab} = \boldsymbol{y}^b \wedge \boldsymbol{w}_J^{ab} \tag{6}$$

where $\boldsymbol{y^b}$ is having boolean coordinates

$$y_k^b = \begin{cases} 1 & \text{if the } k\text{th category won in } ART_b \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

so $|\boldsymbol{x}_{ab}|$ is the value of the weight that predicts the winning ART_b Kth category as a result of the winning ART_a Jth category. The map field performs a vigilance test similarly to the ART_a vigilance test, where if the match function exceeds the map field vigilance parameter $\rho_{ab} \in [0, 1]$

$$\frac{|\boldsymbol{x}_{ab}|}{|\boldsymbol{y}^b|} \geq \rho_{ab} \tag{8}$$

then resonance occurs. This test assures that the prediction of the correct class via the feature vector complies with the label (which is represented by the winning ART_b Kth category). Else, a match tracking procedure is activated for finding a better category in ART_a. In this process, the map field raises ART_a's vigilance parameter ρ_a

$$\rho_a = \frac{|\boldsymbol{I}_a^e \wedge \boldsymbol{w}_J^a|}{|\boldsymbol{I}_a^e|} + \delta, 0 \leq \delta \ll 1. \tag{9}$$

This ensures the Jth category to fail the vigilance test in ART_a and be removed from the competition. The search in ART_a proceeds until an ART_a category that predicts the correct ART_b category is chosen, otherwise a new category is created. The association from an ART_a category to an ART_b category is gained by the following learning rule

$$\boldsymbol{w}_J^{ab,new} = \beta_{ab}(\boldsymbol{w}_J^{ab,old} \wedge \boldsymbol{y}^b) + (1 - \beta_{ab})\boldsymbol{w}_J^{ab,old} \tag{10}$$

which is activated during resonance in the map field. Thus, in fast learning mode ($\beta = 1$) the link from an ART_a Jth category to an ART_b Kth category is permanent. Initial $\boldsymbol{w}_j^{ab}$ weights are assigned to one. In the test and validation phases only ART_a is active, so no vigilance test in the map field occurs. The class prediction is deduced from the map field weights of the winning ART_a category.

3. EXPERIMENTATION

3.1. Methodology

We investigated the fuzzy ARTMAP classifier in discrimination of FISH signals. After image acquisition and analysis we obtained 3008 sample signals represented by twelve features of size, shape, intensity and color [3]. Three classification tasks were experimented: discriminating red signals representing Down syndrome as real or artifact, discriminating green signals representing Patau syndrome as real or artifact and accomplishing both tasks simultaneously (four-class classification problem). The FAM parameters were configured to fast learning ($\beta = 1$, $\beta_{ab} = 1$), the choice parameter α was set to 10^{-6} and ART_a vigilance parameter to 0.5. One troublesome issue of using FAM classifiers is the network inherent sensitivity to the presentation order of training data [2] [4]. In order to overcome this problem we averaged the classification accuracy over 5 experiments

having each a random permutation of the training data. In addition we employed a CV-10 procedure in order to estimate the overall classifier accuracy.

3.2. Training modes

Two modes of training were implemented. In the first mode, called training until completion, the system is repeatedly trained until predicting the labels (almost) perfectly. In the second mode, called training with validation, each training epoch the system accuracy is evaluated on an independent validation set and training is stopped if no further increase of the accuracy is reached. Training with validation is a popular approach in the machine learning community aiming at the elimination of over-fitting in order to enhance generalization [5]. This approach also regards the usage of training until completion as erroneous as it is prone to over-fitting. On the other hand, the developers of FAM defined training until completion as the method of choice and it is used frequently [2]. A comparison of the accuracy on the training, validation (if exist) and test sets, as well as the number of ART_a categories and training epochs for the two training modes and the three previously described classification tasks is given in Table 1. Accuracy on the test set achieved by the training until completion mode outperforms slightly that achieved using the validation set in approximately 1% for all tasks.

3.3. Vigilance sensitivity

We repeated the experiment for the Down syndrome signals and the two training modes for increasing values of ART_a vigilance parameter ρ_a averaging over 20 orderings. Figures 1 and 2 show, respectively, the test, training (and validation) accuracy versus the vigilance parameter for the two training methods. The overfitting caused by the training until completion method can be seen in Figure 2 through the difference in accuracy between training and test and the smaller standard deviation for the training accuracy, as well as by the increase in the number of categories (Table 1) indicating model complexity. Overfitting for the training until completion mode starts later for larger vigilance parameters than for the validation mode of training. It can be also seen that the results are stable up to a certain value of the vigilance parameter.

3.4. A comparison to other classifiers

The fuzzy ARTMAP accuracy is compared to other state-of-the-art classifiers [3] in Table 2. The FAM is comparable to the multi-layer perceptron and inferior to the Bayesian neural network, however requiring less than five epochs to accomplish this accuracy. Both neural networks are superior to the naive Bayesian classifier.

4. DISCUSSION

We have evaluated empirically the fuzzy ARTMAP neural network in the classification of cytogenetic data. Training the FAM until completion produced better test accuracy than training with validation although both training modes overfitted the data. However, overfitting the training set during training until completion had almost no consequences on the test error. This conclusion is in contrast to other machine learning paradigms, and especially neural networks such as a multi layer perceptron, that lose test accuracy due to overfitting. When compared to other state-of-the-art classifiers the fuzzy ARTMAP found to attain similar results to the multi layer perceptron classifier requiring only few training epochs. The inferiority of the FAM to the Bayesian network in classifying the cytogenetic data encourages further investigation into fuzzy ARTMAP .

Acknowledgment This work was supported in part by the Paul Ivanier Center for Robotics and Production Management, Ben-Gurion University, Beer-Sheva, Israel.

5. REFERENCES

[1] H. J. Tanke, R. J. Florijn, J. Wiegant, A. K. Raap, and J. Vrolijk, "CCD microscopy and image analysis of cells and chromosomes stained by fluorescence in situ hybridization," *Histochemical Journal*, vol. 27, pp. 4–14, 1995.

[2] G. Carpenter, S. Grossberg, N. Markuzon, J. Reynolds, and D. Rosen, "Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE transactions on Neural Networks*, vol. 3, pp. 698–713, 1992.

[3] B. Lerner and N. D. Lawrence, "A comparison of state-of-the-art classification techniques with application to cytogenetics," *Neural Computing & Applications*, vol. 10, pp. 39–47, 2001.

[4] I. Dagher, M. Georgiopoulos, G.L. Heileman, and G. Bebis, "An ordering algorithm for pattern presentation in fuzzy artmap that tends to improve generalization performance," *IEEE transactions on neural networks*, vol. 10(4), pp. 768–778, 1999.

[5] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.

Table 1. Fuzzy ARTMAP classification accuracy of FISH signals on the training, validation (if exists) and test sets. Accuracy is given for red and green signals representing Down and Patau syndromes, respectively, and for both syndromes (four-class). Modes 1 and 2 refer to validation and training until completion modes, respectively. Number of ART_a categories and training epochs is included.

classification task	mode	test (%)	train (%)	validation (%)	categories	epochs
four-class	1	80.25 (2.36)	94.36 (3.59)	81.42 (1.94)	155.87 (30.24)	3.47 (1)
	2	81.8 (1.81)	99.33 (0.29)	-	222.45 (10.84)	5.3 (0.71)
red	1	84.16 (3.42)	91.64 (3.94)	85.71 (3.09)	56.18 (19.1)	3.02 (1.14)
	2	85.82 (2.78)	99.39 (0.29)	-	103.74 (7.18)	5.56 (0.72)
green	1	81.34 (3.48)	90.01 (4.08)	81.98 (3.21)	50.66 (15.31)	2.7 (0.88)
	2	82.58 (3.22)	99.46 (0.31)	-	100.28 (7.54)	5.36 (0.74)

Table 2. State-of-the-art classifiers performance compared to fuzzy ARTMAP

Classifier	Test accuracy (%)
Bayesian Neural Network (BNN)	87.14
Neural Network	84.76
fuzzy ARTMAP	84.47
Naive Bayesian Classifier (NBC)	78.02

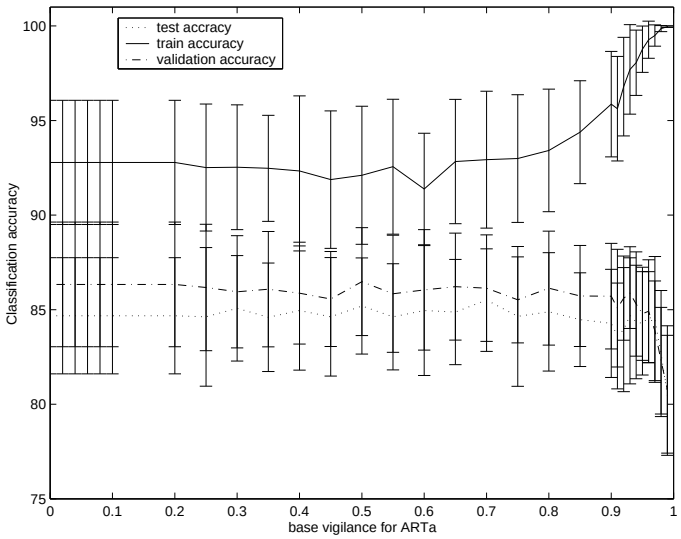

Fig. 1. Accuracy vs. ART_a vigilance parameter for training with validation

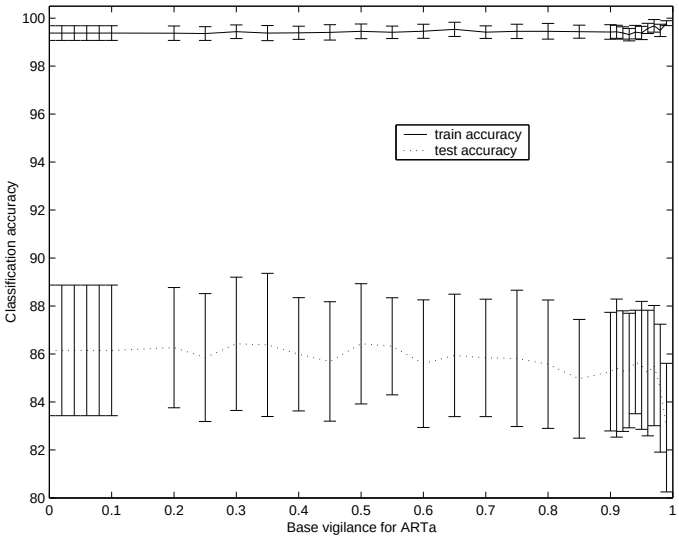

Fig. 2. Accuracy vs. ART_a vigilance parameter for training until completion