Neural Networks 98 (2018) 236–250

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

OnARTMAP: A Fuzzy ARTMAP-based Architecture

Alan L.S. Matias, Ajalmar R. Rocha Neto *

Federal Institute of Ceará, Fortaleza, Ceará, Brazil

ARTICLE INFO

Article history:
Received 21 March 2017
Received in revised form 15 November 2017
Accepted 16 November 2017
Available online 22 December 2017

Keywords:
Adaptive resonance theory
Fuzzy ARTMAP
Category proliferation reduction
Pruning

ABSTRACT

Fuzzy ARTMAP (FAM) copes with the stability–plasticity dilemma by the adaptive resonance theory (ART). Despite such an advantage, Fuzzy ARTMAP suffers from a category proliferation problem, which leads to a high number of categories and a decrease in performance for unseen patterns. Such drawbacks are mainly caused by the overlapping region (noise) between classes. To overcome these drawbacks, we propose a Fuzzy ARTMAP-based architecture robust to noise, named OnARTMAP, for both online and batch learning. Our neural networks (OnARTMAP$_1$ and OnARTMAP$_2$) proposed for batch learning have a two-stage learning process, while our neural network (OnARTMAP$_o$) for online and incremental learning has just a single iterative process. Two new modules are proposed, the overlapping region detection module (ORDM) and another one similar to ART_a, called ART_c. The ORDM finds the overlapping region between categories, while the ART_c computes and stores special categories for overlapping areas. In our architecture proposal, the weights for ordinary categories are estimated from data outside the overlapping area. An alternative to the second stage strategy for batch learning is presented and focuses on improving the generalization performance. On the basis of our achievements, one can infer that OnARTMAP can improve the generalization performance and decrease the number of categories. Our proposals were applied to artificial and real datasets, as well as were compared with several counterparts (Fuzzy ARTMAP, ART-EMAP, μARTMAP, and BARTMAP).

1. Introduction

The Adaptive Resonance Theory (ART) develop by Grossberg (Grossberg, 1976, 1987, 2013) has inspired new architectures of artificial neural networks for both supervised and unsupervised learning. In the ART context, the unsupervised neural networks were the first models to be introduced in the literature. As examples, we can cite ART 1 (Carpenter & Grossberg, 1987), ART 2 (Carpenter & Grossberg, 1990a), ART 3 (Carpenter & Grossberg, 1990b) and the well-known Fuzzy ART (Carpenter, Grossberg, & Rosen, 1991) for unsupervised learning tasks. In a nutshell, ART 1 self-stabilizes recognition codes in response to arbitrary binary input patterns, while ART 2 is able to stably learn to categorize both binary and analog data. On the other hand, Fuzzy ART is an architecture that relies on fuzzy set operators for analog data.

The supervised learning approach based on Grossberg's theory was introduced in the ARTMAP architecture (Carpenter, Grossberg, & Reynolds, 1991). In order to solve this type of task, the ARTMAP uses two ART 1 modules, known as ART_a (the A-side module) and ART_b (the B-side module) for binary data. Along with these modules, there is an inter-ART module responsible for mapping ART_a categories to ART_b categories in a many-to-one approach. As an improvement, Fuzzy ARTMAP was proposed to cope with analog data (Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992).

The Fuzzy ARTMAP can map an arbitrary multidimensional dataset by creating hyper-rectangles for both input pattern and input label in ART_a and ART_b modules, respectively. Some Fuzzy ARTMAP advantages are remarkable, such as fast and stable learning, the need of few epochs to achieve stability, the ability to learn quickly and stably new data without catastrophically forgetting past data, and so on. Despite such advantages, Fuzzy ARTMAP is very sensitive to input pattern presentation order and noise data. This is an important issue concerning the learning process known as the category proliferation problem, since a large number of categories should be included in the module ART_a to represent the input space and its relations to the output space.

The main reason for the category proliferation is the correction of predictive error (in the module ART_b) performed by the match tracking mechanism. Due to the application of this correction process in a very noisy dataset, smaller categories should be created inside larger ones. As a result, the learning process of Fuzzy ARTMAP creates too many small and specialized categories. In fact, the larger the overlapping area between classes in a classification task, the larger is the number of small categories within this region. Therefore, one can see that the category proliferation problem is intensified with the degree of class overlapping.

* Corresponding author.
E-mail addresses: alan.matias@ppgcc.ifce.edu.br (A.L.S. Matias), ajalmar@ifce.edu.br (A.R. Rocha Neto).

https://doi.org/10.1016/j.neunet.2017.11.012
0893-6080/

Several works can be found in the literature to handle the category proliferation problem in Fuzzy ARTMAP. Such works rely mainly on changing the Fuzzy ARTMAP architecture (Carpenter & Markuzon, 1998; Gmez-Snchez, Dimitriadis, Cano-Izquierdo, & Lpez-Coronado, 2000, 2002; Verzi, Heileman, & Georgiopoulos, 2006; Verzi, Heileman, Georgiopoulos, & Healy, 1998; Zhang, Ji, & Zhang, 2014), applying post-processing methods for pruning categories (Carpenter & Tan, 1995; Pourpanah, Lim, & Saleh, 2016) or even replacing the hyper-rectangles with hyper-spheres, gaussians, polytopes, and so on (Amorin, Delgado, & Ameneiro, 2007; Anagnostopoulos & Georgiopoulos, 2000, 2001; Vidgor & Lerner, 2007; Williamson, 1996).

The models that change the Fuzzy ARTMAP architecture include the Boosted ARTMAP (Verzi et al., 2006, 1998), which allows non-zero training error in order to improve overall generalization; the μARTMAP (Gmez-Snchez et al., 2000, 2002), which uses a probabilistic setting to optimize the categories sizes; and Threshold and Posterior Probability FAM (TPPFAM) (Zhang et al., 2014), which performs threshold filtering before a category is committed in training stage, and improves accuracy by combining category information distributed by a dynamic Q-max rule and posterior probability estimated during test. Along with these works, we can also cite the algorithm MT- (Carpenter & Markuzon, 1998), which makes the ART_a search process to find a new category in case of mismatch in the map field easier by decreasing the value of the vigilance parameter (i.e., $\varepsilon \leq 0$; we detail this next section). Similarly, the algorithm MT+ stands for the usual process of raising the vigilance parameter in Fuzzy ARTMAP.

Models that apply post-processing methods for pruning categories are: GA-QFAM (Pourpanah et al., 2016), which uses a Fuzzy ARTMAP classifier with Q-learning for incremental learning, and applies a Genetic Algorithm for rule extraction; and in Carpenter and Tan (1995), Carpenter and Tan investigate different methods for rule extraction on Fuzzy ARTMAP.

Moreover, the models that change the category geometry include Gaussian ARTMAP (Williamson, 1996), which is a synthesis between a Gaussian classifier and the adaptive resonance theory; Hypersphere ARTMAP (Anagnostopoulos & Georgiopoulos, 2000) and Ellipsoid ARTMAP (Anagnostopoulos & Georgiopoulos, 2001), which uses, respectively, hyper-spheres and hyper-ellipsoids for data generalization; Polytope ARTMAP (Amorin et al., 2007), in which categories are irregular polytopes; and Bayesian ARTMAP, which uses a Bayesian framework to improve generalization, while reducing the number of created categories.

By analyzing the previous literature that attempts to solve the category proliferation problem, one cannot find one that aims, at first, to obtain the overlapping region between classes under the ART framework. With this in mind, our work focuses on solving the category proliferation problem by detecting the overlapping area (i.e., noisy region) and then creating ART_a categories placed outside this critical zone. To achieve this, we propose a novel Fuzzy ARTMAP-based architecture that can identify the overlapping regions between classes, if they exist, and exclude this noisy data from the training dataset to prevent the creation of unnecessary categories. This is accomplished by including an additional Fuzzy ARTMAP module, named ART_c, for storing overlapping categories, and a module for overlapping region detection henceforth called the Overlapping Region Detection Module (ORDM). Our proposal is called opposite-to-noise ARTMAP (OnARTMAP), since it is able to learn the data outside the noise area.

Our contributions with this work are: (i) an improved Fuzzy ARTMAP-based architecture; (ii) a new Fuzzy ARTMAP module to detect (obtain) overlapping area (information) between classes; (iii) a novel Fuzzy ARTMAP module for storing overlapping information in special categories; and (iv) a comparative study of our proposal with other related Fuzzy ARTMAP neural networks.

Our paper is organized as follows. In Section 2, Fuzzy ARTMAP architectures are briefly described and, in Section 3, our proposal is presented. In Section 4, we present the results of our simulations for OnARTMAP and its counterparts. In this section, results for artificial and real problems are also included. After that, we present in Section 5 the conclusions and future works.

2. Fuzzy ARTMAP

The Fuzzy ARTMAP architecture has two Fuzzy ART modules, and an additional module to link them, named inter-ART. The Fuzzy ART modules are described below.

2.1. Fuzzy ART

The Fuzzy ART module comprises three layers: F_0, F_1, and F_2. The layer F_0 stands for the current input vector; the layer F_1, that receives both bottom-up input from F_0 (i.e., the output of F_0) and the top-down input from F_2. The layer F_2 represents the active category. The activation (output) for F_0 is the current input vector $\mathbf{I} \in \mathbb{R}^M$ described by $\mathbf{I} = (\mathbf{a}) = (a_1, \ldots, a_i, \ldots, a_M)$, so that $a_i \in [0, 1]$ and the norm $|\mathbf{I}| = \sum_{i=1}^{M} |a_i|$. An alternative representation for the input vector $\mathbf{I} \in \mathbb{R}^{2M}$ is given by the complement coding $\mathbf{I} = (\mathbf{a}, \mathbf{a}^c)$, where $\mathbf{a}^c = 1 - \mathbf{a}$. Note that the norm

$$|\mathbf{I}| = |(\mathbf{a}, \mathbf{a}^c)| = \sum_{i=1}^{M} a_i + (M - \sum_{i=1}^{M} a_i) = M. \tag{1}$$

The outputs for F_1 and F_2 are $\mathbf{x} = (x_1, \ldots, x_M)$ and $\mathbf{y} = (y_1, \ldots, y_N)$, respectively. Moreover, the jth node of F_2 with an adaptive weight vector $\mathbf{w}_j = (w_1, \ldots, w_i, \ldots, w_M)$ is a category representing a training pattern subset. For input vectors with complement coding, the weight vectors $\mathbf{w}_j = (\mathbf{u}_j, \mathbf{v}_j^c)$ are $2M$-dimensional. At the beginning of the training process, each component w_{ji} equals one and, during the process, is monotonically nonincreasing, which lets the learning be stable. As for parameters, the Fuzzy ART has the vigilance parameter $\rho \in [0, 1]$, the choice parameter $\alpha > 0$, and the learning rate $\beta \in [0, 1]$.

For a certain input vector $\mathbf{I}$, the choice function is defined by

$$T_j(\mathbf{I}) = T_j = \frac{|\mathbf{I} \wedge \mathbf{w}_j|}{\alpha + |\mathbf{w}_j|}, \tag{2}$$

where the fuzzy operator AND ($\wedge$) is defined by $\mathbf{x} \wedge \mathbf{y} = min(x_i, y_i)$. Indeed, the **category choice** is given by

$$T_J = \max\{T_j\}_{j=1}^{N}. \tag{3}$$

where J is the index of the chosen category. If more than one T_j is maximal, then the jth category with the lowest index is chosen. In such a situation, $y_J = 1$ and $y_j = 0$, whenever $j \neq J$. The vector $\mathbf{y}$ is the activity for the layer F_2, while the activity for the layer F_1 is given by

$$\mathbf{x} = \begin{cases} \mathbf{I} & \text{if } F_2 \text{ is inactive} \\ \mathbf{I} \wedge \mathbf{w}_J & \text{if } F_2 \text{ is active.} \end{cases} \tag{4}$$

Resonance occurs if the match function, $|\mathbf{I} \wedge \mathbf{w}_J| / |\mathbf{I}|$, of the choice category J meets the **vigilance criterion**

$$\frac{|\mathbf{I} \wedge \mathbf{w}_J|}{|\mathbf{I}|} \geq \rho. \tag{5}$$

In this context, if the vigilance criterion complies with Eq. (5), the choice category $\mathbf{w}_J$ matches and the **updating rule** must be performed, according to

$$\mathbf{w}_J^{new} = \beta(\mathbf{I} \wedge \mathbf{w}_J^{old}) + (1 - \beta)\mathbf{w}_J^{old}. \tag{6}$$