Nested Monte Carlo Search Expression Discovery for the Automated Design of Fuzzy ART Category Choice Functions

Marketa Illetskova*, Islam Elnabarawy*, Leonardo Enzo Brito da Silva*†,
Daniel R. Tauritz*, and Donald C. Wunsch II*
*Missouri University of Science and Technology, Rolla, Missouri, U.S.A.
†CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil
miwdf@mst.edu,elnabarawy@ieee.org,leonardoenzo@ieee.org,dtauritz@acm.org,wunsch@ieee.org

ABSTRACT

While the performance of many neural network and machine learning schemes has been improved through the automated design of various components of their architectures, the automated improvement of Adaptive Resonance Theory (ART) neural networks remains relatively unexplored. Recent work introduced a genetic programming (GP) approach to improve the performance of the Fuzzy ART neural network employing a hyper-heuristic approach to tailor Fuzzy ART's category choice function to specific problems. The GP method showed promising results. However, GP is not the only tool that can be used for automatic improvement. Among other methods, Nested Monte Carlo Search (NMCS) was recently applied to expression discovery and outperformed traditional evolutionary approaches by finding better solutions in fewer evaluations. This work applies NMCS to the discovery of new Fuzzy ART category choice functions targeted to specific problems with results demonstrating its ability to find better performing Fuzzy ART networks than the GP approach.

CCS CONCEPTS

• **Computing methodologies → Heuristic function construction**; *Genetic programming*; Game tree search; Neural networks;

KEYWORDS

genetic programming, neural networks, metaheuristics, algorithm engineering, empirical study

ACM Reference Format:
Marketa Illetskova*, Islam Elnabarawy*, Leonardo Enzo Brito da Silva*†, Daniel R. Tauritz*, and Donald C. Wunsch II*. 2019. Nested Monte Carlo Search Expression Discovery for the Automated Design of Fuzzy ART Category Choice Functions. In *Genetic and Evolutionary Computation Conference Companion (GECCO '19 Companion), July 13–17, 2019, Prague, Czech Republic.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3319619.3322050

GECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-6748-6/19/07…$15.00
https://doi.org/10.1145/3319619.3322050

1 BACKGROUND AND OBJECTIVES

Adaptive Resonance Theory (ART) is a biologically-plausible theory that models how the brain can autonomously learn to categorize, recognize, and predict objects and events in a dynamic and changing environment [1]. There are various ART-inspired neural network models in the literature. However, as with many other neural network models, there may exist optimal ART design parameters that are yet to be discovered.

Numerous approaches have been introduced that automatically design neural networks; yet, the automatic design of ART and ART-based networks remains relatively unexplored. Previous work evolved and trained parameters and weights in ART based networks [7, 8, 12, 14]. Recently, a novel hyper-heuristic approach was introduced that employed genetic programming (GP) to evolve the category choice function (CCF) in Fuzzy ART [2] for specific problem classes [6]. The GP approach showed promising results leading to improved Fuzzy ART performance. While GP is the most commonly used meta-heuristic applied to automatic design and improvement of algorithms, the exploration of the algorithm space may well be addressed with other search methods. This work pursues the Nested Monte Carlo Search Expression Discovery (NMCS-ED) approach to search for new Fuzzy ART CCFs.

The NMCS-ED, an alternative to other hyper-heuristic methods, employs a method inspired by NMCS to find expressions [3, 4]. At the lowest level, NMCS-ED randomly samples the expression and applies nested searches at higher levels [3, 4]. It maintains a good balance between exploration and exploitation, and inherently has a natural restart strategy [3, 4]. One of the challenges related to GP is parameter optimization. GP algorithms often take many parameters such as population size, and mutation probability. In contrast, other than the level of the search, NMCS-ED in its basic form only takes a single parameter: the maximum number of nodes (n) that an expression in a tree form can have. The tree is expanded during the iterations of the algorithm until the tree is full or reaches n nodes. Since its introduction, NMCS-ED was applied to various problems and it was shown that it has the capability of producing equally good, or better, solutions in less evaluations than GP [3–5, 9–11].

Fuzzy ART is an unsupervised learning network consisting of an input representation layer and a category representation layer [2]. This study uses NMCS-ED to automatically design the Fuzzy ART CCF in Eq. (1) that was previously evolved using GP [6]. The CCF measures the activation of a category with weight vector w_j in response to an input pattern vector A, where α is a scalar algorithm parameter known as the choice parameter.

Table 1: Statistical test comparison of Adjusted Rand-Index results for the NMCS-ED and GP approaches and the baseline. With $p = 0.05$ as the cutoff for statistical significance, values underlined significantly outperformed the baseline, values in bold significantly outperformed the GP approach.

dataset	*approach*	*avg*	*stdev*	*p values to baseline*	*p values to GP*
Iris	baseline	0.526977	0.042583	–	–
Iris	GP	0.526695	0.044548	9.886139e-01	–
Iris	NMCS-ED 16	0.545040	0.040105	3.417641e-01	3.459611e-01
Iris	NMCS-ED 40	0.554020	0.035245	1.392459e-01	1.455892e-01
Iris	NMCS-ED 100	0.563740	0.040259	6.273904e-02	6.679941e-02
Wine	baseline	0.133080	0.051846	–	–
Wine	GP	<u>0.263992</u>	0.058357	4.838685e-05	–
Wine	NMCS-ED 16	0.240686	0.071779	1.191519e-03	4.360158e-01
Wine	NMCS-ED 40	<u>**0.317218**</u>	0.050713	2.327790e-07	4.302810e-02
Wine	NMCS-ED 100	<u>**0.459520**</u>	0.053130	4.546132e-11	3.295069e-07
Glass	baseline	0.148816	0.030336	–	–
Glass	GP	0.139372	0.028285	4.807511e-01	–
Glass	NMCS-ED 16	<u>**0.175204**</u>	0.024102	4.506389e-02	6.904851e-03
Glass	NMCS-ED 40	<u>**0.178915**</u>	0.023738	2.369253e-02	3.289170e-03
Glass	NMCS-ED 100	<u>**0.197412**</u>	0.018720	4.207209e-04	3.847010e-05

$$T_j = \frac{|A \wedge w_j|}{\alpha + |w_j|} \tag{1}$$

The NMCS-ED algorithm introduced in [3] uses a stack based approach. To automatically generate new CCFs, we represent them as parse trees with terminal nodes A, w_j, and α, and non-terminal nodes for addition, subtraction, multiplication, division, Fuzzy AND, Fuzzy OR, and L^1 Norm operators [6]. To ensure the compatibility of the arguments and operators, and to generate valid expressions, we modified the NMCS-ED algorithm from [3] to use a tree-based approach capable of generating strongly-typed expressions.

2 EXPERIMENTS AND CONCLUSIONS

In our experiments, we capped the number of evaluations at 10,000 since that is approximately the amount of evaluations that was employed in the GP approach. For a fair comparison, we used the same method of evaluation of the newly generated CCFs as was used in [6], and used the same datasets for training and testing - the Iris, Wine, and Glass datasets from the UCI Machine Learning Repository [13]. Furthermore, we used the same values for the Fuzzy ART parameters α, β, and ρ as those found for the GP approach for each of the datasets [6]. Since n is the unknown parameter in NMCS-ED that is to be configured, we experimented with three different values - 100 being an approximation of tree height selected in [6] and 40 and 16 values selected to evaluate the quality of smaller expressions, since smaller expressions require shorter run times.

Table 1 shows the baseline values for standard Fuzzy ART and averages of ten runs of the GP approach on each of the used datasets [6]. Further, it shows the averages of the ten runs of the NMCS-ED approach with $n \in \{16, 40, 100\}$, and t-student statistical test results comparing the two approaches and the baselines. We used a two-tailed t-student statistical test with p=0.05 to determine the statistical significance (same statistical test setting as in [6]).

Per Table 1, on the Iris dataset the average performance of the Fuzzy ART networks with CCFs found using NMCS-ED was statistically similar to the performance of networks with CCFs found using GP. The averages of the performances of the expressions found using NMCS-ED improved with larger n. On the Wine dataset, for $n = 16$, NMCS-ED performed statistically similar to GP. On the other hand, for $n = 40, 100$, NMCS-ED statistically significantly outperformed GP. On the Glass dataset, NMCS-ED statistically significantly outperformed GP for all n tested. Our experiments were capped with maximum allowed run time of seven days. Seven out of the ten experiments on the Glass dataset with $n = 100$ did not finish in that time, yet, several of those experiments found solutions significantly better than the baseline, and even significantly better than the ones found by the GP approach. Furthermore, the overall best performing Fuzzy ART's CCFs for each of the datasets were found using the NMCS-ED approach.

This work examined the use of the NMCS-ED algorithm applied to the automated design of CCFs for the Fuzzy ART neural network. The fitness function for the optimization was defined as the Adjusted Rand Index; thus, the category choice functions were tailored in a supervised manner. Results show that, for the benchmark datasets experimented with, the networks embedded with the functions automatically redesigned by the NMCS yielded statistically comparable or superior performance to the ones constructed by the GP as well as the canonical Fuzzy ART. Thus, the NMCS-ED based hyper-heuristic may serve as a good alternative to GP in automatic enhancement and design of ART-based systems.

REFERENCES

[1] G. A. Carpenter and S. Grossberg. 1988. The ART of Adaptive Pattern Recognition by a Self-organizing Neural Network. *IEEE Computer* 21, 3 (Mar 1988), 77–88.
[2] G. A. Carpenter, S. Grossberg, and D. B. Rosen. 1991. Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System. *Neural Netw.* 4, 6 (Nov 1991), 759–771.
[3] T. Cazenave. 2010. Nested Monte-Carlo Expression Discovery. In *ECAI*. 1057–1058.
[4] T. Cazenave. 2013. Monte-Carlo Expression Discovery. *International Journal on Artificial Intelligence Tools* 22, 01 (2013).
[5] T. Cazenave and S. Ben Hamida. 2015. Forecasting Financial Volatility Using Nested Monte Carlo Expression Discovery. In *2015 IEEE Symposium Series on Computational Intelligence*. 726–733.
[6] I. Elnabarawy, D. R. Tauritz, and D. C. Wunsch. 2017. Evolutionary Computation for the Automated Design of Category Functions for Fuzzy ART: An Initial Exploration. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '17)*. ACM, New York, NY, USA, 1133–1140.
[7] A. Kaylani, M. Georgiopoulos, M. Mollaghasemi, and G.C. Anagnostopoulos. 2009. AG-ART: An Adaptive Approach to Evolving ART Architectures. *Neurocomputing* 72, 10 (2009), 2079 – 2092.
[8] A. Kaylani, M. Georgiopoulos, M. Mollaghasemi, G. C. Anagnostopoulos, C. Sentelle, and M. Zhong. 2010. An Adaptive Multiobjective Approach to Evolving ART Architectures. *IEEE Transactions on Neural Networks* 21, 4 (Apr 2010), 529–550.
[9] T. Křen, J. Moudřík, and R. Neruda. 2017. Combining Top-down and Bottom-up Approaches for Automated Discovery of Typed Programs. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. 1–8.
[10] G. Kronberger and M. Kommenda. 2013. Evolution of Covariance Functions for Gaussian Process Regression using Genetic Programming. *ArXiv e-prints* (May 2013).
[11] R. Lee, M. J. Kochenderfer, O. J. Mengshoel, and J. Silbermann. 2017. Interpretable Categorization of Heterogeneous Time Series Data. *ArXiv e-prints* (Aug 2017).
[12] R. Li, T. R. Mersch, O. X. Wen, A. Kaylani, and G. C. Anagnostopoulos. 2010. Multi-objective Memetic Evolution of ART-based Classifiers. In *IEEE Congress on Evolutionary Computation*. 1–8.
[13] M. Lichman. 2013. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml.
[14] S. C. Tan, J. Watada, Z. Ibrahim, and M. Khalid. 2015. Evolutionary Fuzzy ARTMAP Neural Networks for Classification of Semiconductor Defects. *IEEE Transactions on Neural Networks and Learning Systems* 26, 5 (May 2015), 933–950.

Sponsored by the Army Research Laboratory and Lifelong Learning Machines program from DARPA/MTO under Cooperative Agreement Number W911NF-18-2-0260.