November 9, 2024

Dear Editors,

We are pleased to submit our manuscript *Multi-Objective Linear Ensembles for Robust and Sparse Training of Few-Bit Neural Networks* for publication in INFORMS Journal on Computing.

In this work we propose new methods to improve the Mixed Integer Linear Programming (MILP) training of Binarized Neural Networks (BNNs), whose values are restricted to $\{\pm 1, 0\}$, and Integer Neural Networks (INNs), whose values lie in the integer range $-P, \ldots, P$.

In particular, our contribution is a multi-objective ensemble approach. We train a single NN for each possible pair of classes and apply a majority voting scheme to predict the final output. Moreover, each network is trained with a multi-objective function that results in the training of robust sparsified networks whose output is not affected by small perturbations on the input, and whose number of active weights is as small as possible. We provide computational evidence that our method achieves its goal compared to the state of the art in the literature.

A preliminary version of parts of this paper was presented at the LION 2023 Conference (Nice, France, June 2023) and will appear as a post-conference proceedings in LNCS. This submitted manuscript extends the conference paper in the following important ways:

Motivations. This paper better motivates and situates our approach in the state of the art. A new section on Related Works is added, in which we extend the original bibliography and motivate our approach with respect to the existing literature.

Integer Neural Networks. Our ensemble approach and lexicographic multi-objective function are developed for Integer Neural Networks (INNs), generalizing the case of Binarized Neural Networks (BNNs) and solving MILPs in a much larger research space.

Empirical results. This paper presents more extensive and improved empirical results. All the experiments presented in the original proceeding paper are performed with a newer version of Gurobi over a larger number of instances. Moreover, new empirical results are presented. In particular, an ablation study regarding the different steps of the lexicographic multi-objective function is conducted. In addition, a comparison between various INNs is performed over the MNIST dataset and a new dataset: the Heart Disease dataset.

New analysis. The new experiments presented in this paper lead to a novel analysis of the distribution of INN weights. Empirical results on the MNIST dataset and the Heart Disease dataset show how extremal-valued distributions are preferred over uniform distributions.

Dr Neil Yorke-Smith played a fundamental role in driving the developments outlined above. In fact, he invited the first authors Ambrogio Maria Bernardelli and Simone

Milanesi to spend a one-month stay at the TU Delft, where the collaborative effort enabled the two Ph.D. students to harness his extensive expertise, leading to significant enhancements in both the theoretical framework and experimental design put forth in the preceding LION paper. Also, he contributed in both writing and reviewing. For these reasons, we decided to invite him as a co-author for this paper.

We believe that our paper is a perfect fit for IJOC, in particular for the area *Design & Analysis of Algorithms – Discrete*. We propose a MILP approach to a Neural Network training problem and the interaction between Optimization models and Machine Learning / Decision-Making problems is widely covered from IJOC in the recent published [**?**, **?**, **?**].

Thank you for your time in considering our manuscript.

Yours faithfully,

Ambrogio Maria Bernardelli, Simone Milanesi, Stefano Gualandi
Università degli Studi di Pavia, Dipartimento di Matematica "F. Casorati", Italy

Hoong Chuin Lau
Singapore Management University, Singapore

Neil Yorke-Smith
Delft University of Technology, Netherlands