Binary Quantization Analysis of Neural Networks Weights on MNIST Dataset

Zoran H. Peric[1], Bojan D. Denic[1], Milan S. Savic[2], Nikola J. Vucic[1, *], Nikola B. Simic[3]

[1]*Faculty of Electronic Engineering, University of Nis,*
Aleksandra Medvedeva 14, 18000 Nis, Serbia
[2]*Faculty of Sciences and Mathematics, University of Pristina,*
Ive Lole Ribara 29, 38220 Kosovska Mitrovica, Serbia
[3]*Faculty of Technical Sciences, University of Novi Sad,*
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
nikola.vucic@elfak.ni.ac.rs

Abstract—This paper considers the design of a binary scalar quantizer of Laplacian source and its application in compressed neural networks. The quantizer performance is investigated in a wide dynamic range of data variances, and for that purpose, we derive novel closed-form expressions. Moreover, we propose two selection criteria for the variance range of interest. Binary quantizers are further implemented for compressing neural network weights and its performance is analysed for a simple classification task. Good matching between theory and experiment is observed and a great possibility for implementation is indicated.

Index Terms—Image classification; Multilayer perceptron; Neural network; Quantization; Source coding.