Title: Simulating parallel scalable Probabilistic Neural Networks via Exemplar Selection and EM in a Ring Pipeline

Authors: Yiannis Kokkinos, Konstantinos G. Margaritis

PII: S1877-7503(17)30805-0
DOI: http://dx.doi.org/doi:10.1016/j.jocs.2017.07.008
Reference: JOCS 723

To appear in:

Received date: 9-12-2016
Revised date: 24-5-2017
Accepted date: 10-7-2017

Please cite this article as: Yiannis Kokkinos, Konstantinos G.Margaritis, Simulating parallel scalable Probabilistic Neural Networks via Exemplar Selection and EM in a Ring Pipeline, Journal of Computational Sciencehttp://dx.doi.org/10.1016/j.jocs.2017.07.008

Simulating parallel scalable Probabilistic Neural Networks via Exemplar Selection and EM in a Ring Pipeline

Yiannis Kokkinos and Konstantinos G. Margaritis

Parallel and Distributed Processing Laboratory, Department of Applied Informatics, University of Macedonia, 156 Egnatia str., P.O. Box 1591, 54006, Thessaloniki, Greece

HIGHLIGHTS

- Simulating a new large scale parallel Probabilistic Neural Network (PNN) construction via exemplar selection and expectation maximization
- Application of proposed Kernel Gradient Subtractive Clustering (KG-SC) exemplar selection for the PNN kernels of every class without any user-defined parameters
- Application of a new ring pipeline topology for all parallel algorithms
- Data and PNN neurons are distributed across the processors
- Allows training a large scale distributed model on a large scale distributed dataset

Abstract. We present a new scalable Probabilistic Neural Network (PNN) construction method suitable for data-neuron parallelism in a ring pipeline parallel topology that allows training a large scale distributed model on a large scale distributed dataset. First the recently proposed Kernel Gradient Subtractive Clustering (KG-SC) automatically selects representative exemplar centers and their number for the PNN kernels. Then Expectation Maximization (EM) refines the PNN parameters. Experimental simulations compare the proposed solution accuracy and performance with PNNs produced from other state-of-the-art *k*-center clustering algorithms. The parallel and distributed implementations produce speedups close to linear on increasing the number of processors and the dataset size.

Keywords. Parallel processing, Probabilistic Neural Networks, Kernel Gradient, Subtractive Clustering, Expectation Maximization.

1 Introduction

Intelligent algorithms and scalable data-driven modelling combined with parallel computing emerge as vital parts of computational science [1]. The reasons lie in the