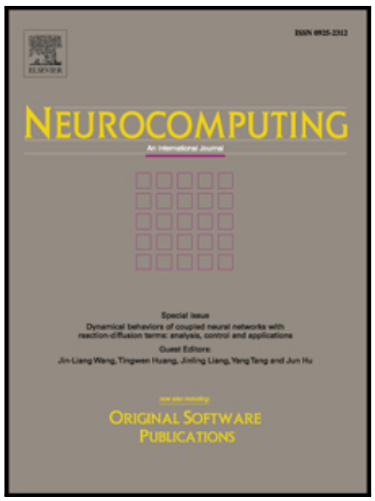

High accuracy FPGA activation function implementation for neural networks

Zbigniew Hajduk

PII: S0925-2312(17)30559-3
DOI: 10.1016/j.neucom.2017.03.044
Reference: NEUCOM 18272

To appear in: *Neurocomputing*

Received date: 8 August 2016
Revised date: 26 February 2017
Accepted date: 17 March 2017

Please cite this article as: Zbigniew Hajduk , High accuracy FPGA activation function implementation for neural networks, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.03.044

High accuracy FPGA activation function implementation for neural networks

Zbigniew Hajduk

Rzeszów University of Technology, ul. Powstańców Warszawy 12, 35-959 Rzeszów, Poland.

e-mail address: zhajduk@kia.prz.edu.pl

Abstract This letter shortly presents an FPGA implementation method of the hyperbolic tangent and sigmoid activation functions for artificial neural networks. A kind of a direct implementation of the functions is proposed. The implementation results show that the obtained accuracy of the method is relatively high compared to other published solutions.

1. Introduction

The most important, expensive and hard to implement part of any hardware realization of artificial neural networks (ANNs) is a neuron's non-linear activation function [1]. A number of papers consider different field programmable gate arrays (FPGAs) implementation methods of the most commonly used hyperbolic tangent and sigmoid activation functions. For example, the piece-wise linear (PWL) interpolation is presented in [2, 3]. A look-up table (LUT) with a linear interpolation between the LUT's points is exercised in [4]. A study of polynomial approximation exercising Lagrange, Chebyshev and least square method is presented in [5]. The usage of the coordinate rotation digital computer (CORDIC) algorithm is featured in [1]. In this letter a kind of a direct implementation of the hyperbolic tangent and sigmoid activation functions is proposed. The implementation difficulty in this approach is shifted to the approximation of the exponent function. Applying a LUT with either the McLaurin series or Padé polynomials is proposed for the approximation realization. Instead of fixed point arithmetic, a single precision floating point (FP) arithmetic is used. The main goal of the proposed implementation method is to attain as high of an accuracy as possible.

2. FPGA implementation