Sum of Step Approximation of a Novel Nonlinear Activation Function

Venkatappreddy.P, M. Deepthi

Abstract—In this manuscript, we propose sum of steps approximation of a novel nonlinear activation function tunable ReLU for VLSI architecture implementation of neural networks. The characteristics of the proposed activation function depend on a tunable parameter and input data set values. Also, we propose a linear-in-the-parameter model for the proposed activation function using an even mirror Fourier nonlinear filter. Finally, simulation results are presented to show performance of the proposed activation function on various data sets and observe its superiority against other activation functions.

Keywords—Activation function, perceptron, tunable ReLU, deep neural network, EMFN filter.

I. INTRODUCTION

Single and multi-layer neural networks are computational training architectures that propagate data through linear and nonlinear operators. Nonlinear operators are also called activation functions that have been found to be very effective when the network has more than one layer. In [1], LeCun introduces deep learning networks that consist of more than five layers. As the number of layers in the deep neural networks (DNN) increase, its performance gets better at the cost of computational complexity. In [2], authors discuss interpretation of DNN model and its applications on various datasets for industrial applications. In [3], authors discuss advantages of deep learning for classification and regression problems in high dimension. DNNs are used in various applications such as monaural speech enhancement in various back ground noise conditions [4], image quality assessment [5], natural language understanding [6], speech recognition [7], and others.

In the literature, various nonlinear activation functions have been studied viz., sigmoid, $tanh$ and rectified linear unit (ReLU). It is observed that an activation function [8] which gives better classification accuracy for one dataset may not give better classification accuracy for other in industrial applications. Sigmoid and $tanh$ are two smoothly varying nonlinear activation functions however they cannot be effectively used in DNN architecture due to gradient vanishing problem. Currently, ReLU is a popular activation function since it eases the gradient vanishing problem. It was proposed for Boltzmann machines and was used in the neural networks by Glorot et al., [9]. Sigmoid and $tanh$ are continuous differential functions with point-wise contractive nature. Since Lipschitz constants for these activation functions are 0.25 and 0.5 respectively [10]. Whereas ReLU is not differentiable at zero and its Lipschitz constant is 1. Most recently other nonlinear activation functions exponential linear unit (ELU) and scaled exponential linear units (SELU) were introduced in [11, 12] that speeds up training in neural network and performs better than ReLU. A closed form approximation for sigmoid and $tanh$ functions by sum of steps for hardware implementation are discussed in [13]. An even mirror Fourier nonlinear (EMFN) filter [14] is used to obtain a linear-in-the-parameter (LIP) model for any strong nonlinear characteristic activation function in DNN for better mathematical tractability.

A small percentage increment in the training accuracy requires an additional number of layers in the network which increases computational complexity. In this manuscript, we propose a novel nonlinear activation function tunable ReLU for characterization of neural networks mathematically and training networks efficiently. First, we study the behavior of the proposed activation function against its tunable parameter. We also approximate tunable ReLU by sum of steps representation for VLSI implementation. Further, we discuss a LIP model for the proposed activation function using an EMFN filter. This LIP model is used to completely represent neural networks with linear and LIP operators.

This manuscript is organized as follows. In Section II, we discuss various existing activation functions and an EMFN filter basis to approximate strong nonlinear systems as preliminaries. In Section III, we discuss the definition and a polynomial approximation of tunable ReLU. Also, we discuss an approximation of tunable ReLU using sum of steps and its LIP model using an EMFN filter. In Section IV, we present the performance results of the proposed activation function. Conclusions follow in Section V.

II. PRELIMINARIES

In this section, we present a brief summary of two concepts that are used in this manuscript. We first talk of definition of an activation function and representations of various activation functions.

Definition: A function is called an activation function $\rho(x) : \mathbb{R} \rightarrow \mathbb{R}$ that is differentiable everywhere mostly [15].

Some of the activation functions used in neural networks to train the datasets are:

1. Sigmoid ($\sigma(x)$): $\frac{1}{1 + e^{-x}}$
**SUM OF STEP APPROXIMATION OF A NOVEL NON LINEAR ACTIVATION FUNCTION**

- Hyperbolic Tangent: $\tanh(x)$
- ReLU: $\max(0, x)$
- Exponential linear unit (ELU):
  $g(x) = \begin{cases} \alpha x e^x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$
  where $\alpha$ is a parameter that controls the value to which an ELU saturates for negative net inputs. Scaled exponential linear units (SELU) is introduced in [12] as an activation function and it is defined as
  $g(x) = \begin{cases} \lambda x e^x & \text{if } x > 0 \\ \lambda x (e^x - 1) & \text{if } x \leq 0 \end{cases}$

  Where, $\lambda$ is a parameter for controlling the mean of the activation function.

(a) Even Mirror Fourier nonlinear filter

In this manuscript, an EMFN filter is used to obtain a LIP model for the proposed activation function. A LIP model is used to characterize nonlinear activation functions in NN, DNN and LSTM architectures mathematically. The properties of Fourier series such as an even mirror repetition and periodicity are used to obtain an EMFN filter basis [14].

Let $h(x)$ is a periodic signal with period 4 and an even mirror repetition property such that

$h(x + 4) = h(x)$

$h(x + 4) = h(x - 4)$

Then the Fourier basis of $h(x)$ that satisfy properties of an Algebra is of the form [16, 17],

$$\{1, \sin \left(\frac{\pi n}{2}\right), \cos \left(\frac{2\pi n}{2}\right), \sin \left(\frac{3\pi n}{2}\right), \ldots \} \quad (2)$$

The above basis functions are useful only when the input sample values are in the range of $[-1, 1]$. The EMFN filter basis is used to approximate any continuous function based on Stone-Weierstrass theorem which states that:

"Given $h : [a, b] \rightarrow R$ continuous and an arbitrary $\varepsilon > 0$ there exists an algebraic polynomial $k$ such that $|h(x) - k(x)| \leq \varepsilon$, $\forall x \in [a, b]"$ [18].

**III. A NOVEL TUNABLE RELU AS ACTIVATION FUNCTION**

In this section, we introduce the tunable ReLU as a novel nonlinear activation function to classify datasets for various industrial applications and mathematical understanding of NN, DNN and LSTM architectures. The proposed activation function is differentiable at every point and eases the gradient vanishing problem. The proposed tunable ReLU representation is obtained by considering the definition of a ReLU activation function. It is denoted as $R(x)$ and represented as

$R(x) = \max(0, x) \ \forall \ x \in R \quad (3)$

ReLU activation function in (3) can be written in terms of signum function as

$R(x) = \frac{1}{2} (1 + \text{sign}(x)) \quad (4)$

where, $\text{sign}(.)$ represents the signum function defined as

$$\text{sign}(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases}$$

The signum function is approximated using tanh function as

$$\text{sign}(x) \approx \tanh(kx) \text{ for } |kx| > 0 \quad (5)$$

where, $k \in R^+$. By substituting (5) in (4), we obtain the proposed tunable ReLU $\rho(x)$ as

$$\rho(x) = \frac{1}{2} (1 + \tanh(kx)) \text{ for } |kx| > 0 \quad (6)$$

$$= \frac{xe^{kx}}{2} \text{sech}(kx) \text{ for } |kx| > 0 \quad (7)$$

We use the equality $\text{sech}(x) = \frac{2}{e^x + e^{-x}}$ in (7) to obtain another form of the tunable ReLU based on a parameter value $k$ which is differentiable at every point as

$$\rho(x) = \frac{1}{1+e^{-2kx}} \text{ for } |kx| > 0 \quad (8)$$

The tunable ReLU $\rho(x)$ and its derivative $\rho'(x)$ are also expressed using sigmoid functions as

$$\rho(x) = x \sigma(2kx) \text{ for } |kx| > 0 \quad (9)$$

Where, $\sigma(2kx)$ is defined as

$$\sigma(2kx) = \begin{cases} 1 & \text{if } kx > 0 \\ 0 & \text{if } kx < 0 \\ \frac{1}{2} & \text{if } kx = 0 \end{cases} \quad (10)$$

using this $\rho'(x)$ is found as

$$\rho'(x) = 2kx \sigma'(2kx) + \sigma(2kx) \text{ for } |kx| > 0 \quad (11)$$

where, $\sigma'(x)$ is the derivative of $\sigma(x)$. We present another form of activation function as

$$\rho(x) = \begin{cases} \frac{x}{k} & \text{for } kx > 0 \\ 0 & \text{for } kx < 0 \\ \frac{x}{2} & \text{for } kx = 0 \end{cases} \quad (12)$$

The proposed activation function is also represented as

$$\rho(x) = xe^{k(x-|x|)} \text{ for } |kx| > 0 \quad (13)$$

where $|.|$ denotes modulus operator. By using the $e^{x} = 2^{x \ln 2}$ and the value of $\ln 2 = 1.4428$, we obtain another form of the proposed activation function as

$$\rho(x) = x 2^{\frac{1}{2} \cdot 4428 \cdot (x-|x|)} \text{ for } |kx| > 0 \quad (14)$$

The proposed activation function is represented using a modulus operator and powers of '2' as shown in Fig. 1. Where, the multiplication operator is denoted as 'X'. The linearized functions $l_i$ and $l_o$ of sigmoid and tanh functions are obtained using Taylor series expansion about zero. The linear approximation of sigmoid function is given as

$$l_i = 0.25x + 0.5$$

Fig. 1: Representation of tunable ReLU.
A. A close approximation of the tunable ReLU by sum of steps

In this subsection, we find an efficient method to approximate the proposed tunable ReLU function using sum of steps to aid hardware implementation. We consider following representation for the proposed activation function.

\[ \rho(x) = \frac{x}{1 + e^{-2kx}} \quad \forall \ x \in R \quad (17) \]

From this function to obtain sum of steps representation we use the following two steps:

(i) We replace \( e \) by 2 since computers use binary number system.

(ii) We replace \( x \) by \( n^{1} \) to obtain fixed point representation of tunable ReLU as

\[ \rho(n) = \sum_{l=0}^{n-1} \sum_{i=1}^{\frac{n}{2i}} \frac{n}{2i} e^{-2kln_{2}} \quad \forall \ n \in Z \quad (18) \]
SUM OF STEP APPROXIMATION OF A NOVEL NON LINEAR ACTIVATION FUNCTION

[\because kx is a large value, we have approximated (2\ln2)e^{kx} by kn]. For larger values of kn the above approximation becomes sum of steps approximation for ReLU function as shown in Fig. 4. The quantization error \( \delta = \rho(x) - \rho(n) \) lies between \([0 1]\) as shown in Fig. 5. By using two’s complement form of \(n\) and rounding, we reduce this error.

B. A linear-in-the-parameter model for the proposed tunable ReLU

In this subsection, a LIP model for the tunable ReLU function in (8) is obtained using an even mirror Fourier nonlinear filter. A time invariant, nonlinear, causal, memoryless system is approximated using an even mirror Fourier nonlinear filter basis. A time invariant, nonlinear, causal, memoryless system is approximated using an even mirror Fourier nonlinear filter basis. A time invariant, nonlinear, causal, memoryless system is approximated using an even mirror Fourier nonlinear filter basis.

The representation of the proposed tunable ReLU using an even mirror Fourier nonlinear filter basis.

IV. PERFORMANCE ANALYSIS & RESULTS

The simulation results show that the proposed activation function tunable ReLU gives better training accuracy in classifying data sets than other activation functions such as sigmoid, \( \tanh \), ReLU and ELU. In the proposed activation function, we consider \( k = 0.01 \) for which the learning rate of network is faster than other activation functions as shown in Fig. 8. The training accuracies using various activation functions for MNIST data set is listed in Table I. From Fig. 9, we observe that the training accuracy of the network using the proposed activation function increases as the number epochs increase.
We have studied the Fourteenth International Conference on Artificial Neural Networks. In Advances in neural information processing systems, pp. 3104-3112. 2014.


A. Carini, G. L. Sicuranza, "Recursive even mirror Fourier nonlinear filter for better mathematical tractability. Finally, the simulation results have shown that the proposed activation function gives better classification accuracy.

V. CONCLUSIONS

In this paper, we have proposed tunable ReLU as a novel nonlinear activation function. We have obtained a polynomial approximation of this activation function. An approximation of the tunable ReLU has obtained using sum of steps for hardware implementations. We have studied characteristics of the proposed activation function, it acts as a ReLU for larger tunable parameters and input data values. We have obtained a linear-in-the-parameter model for this function using an even mirror Fourier nonlinear filter for better mathematical tractability. Finally, the simulation results have shown that the proposed activation function gives better classification accuracy.

REFERENCES