IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 4 Issue 6, June 2017
ISSN (Online) 2348 – 7968 | Impact Factor (2016) – 5.264
www.ijiset.com

A Note on the Soboleva' Modified Hyperbolic Tangent Activation Function

Angel Golev, Anton Iliev and Nikolay Kyurkchiev

Faculty of Mathematics and Informatics, University of Plovdiv Paisii Hilendarski,
24, Tzar Asen Str., 4000 Plovdiv, Bulgaria, e-mails: angel.golev@fmi-plovdiv.org, aii@uni-plovdiv.bg, nkyurk@uni-plovdiv.bg

Abstract

In this note we construct a family of parametric Soboleva' modified hyperbolic tangent activation function (PSMHTAF).
We prove upper and lower estimates for the Hausdorff approximation of the sign function by means of this family. Numerical examples, illustrating our results are given.
Keywords: *Parametric Soboleva' modified hyperbolic tangent activation function (PSMHTAF), Sign function, Hausdorff distance, Upper and lower bounds.*

1. Introduction

Sigmoidal functions (also known as "activation functions") find multiple applications to neural networks [8]–[18].

The modified hyperbolic tangent is a special S-shaped function constructed on the basis of the hyperbolic tangent function, which is expressed in terms of the exponent.

The function looks like [1]–[3]:

$$m(t;a,b,c,d) = \frac{e^{at} - e^{-bt}}{e^{ct} + e^{-dt}}.$$

The function $m(t;a,b,c,d)$ first proposed by Soboleva E. as a utility function of particular criteria for multi-criteria optimization and multifactorial decision-making find application to approximate the current-voltage characteristics of light-emitting diodes [4].

We study the distance between the sign function and a special class of activation functions, so-called parametric Soboleva' modified hyperbolic tangent activation function (PSMHTAF).

The distance is measured in Hausdorff sense, which is natural in a situation when a sign function is involved. Precise upper and lower bounds for the Hausdorff distance are reported.

Any neural net element computes a linear combination of its input signals, and uses a logistic function to produce the result; often called "activation" function [19]– [20].

2. Preliminaries

The following are common examples of activation functions:

a) logistic

$$\varphi_1(t) = \frac{1}{1+e^{-t}}; \tag{1}$$

b) Parametric Hyperbolic Tangent Activation (PHTA) function

$$\varphi_2(t) = \frac{e^{\beta t} - e^{-\beta t}}{e^{\beta t} + e^{-\beta t}} = 1 - \frac{2e^{-\beta t}}{e^{\beta t} + e^{-\beta t}}, \quad t \in \mathbb{R}, \ \beta \geq 1; \tag{2}$$

c) Parametric Half Hyperbolic Tangent Activation (PHHTA) function

$$\varphi_3(t) = \frac{1-e^{-\beta t}}{1+e^{-\beta t}}, \ t \in \mathbb{R}, \ \beta \geq 1. \tag{3}$$

d) Parametric Fibonacci hyperbolic tangent activation function (FHTAF) [38] based on the Fibonacci hyperbolic tangent function [7]

$$\varphi_4(t) = \frac{\Psi^{\beta t} - \Psi^{-\beta t}}{\Psi^{\beta t} + \Psi^{-\beta t}}, \ t \in \mathbb{R}, \ \beta \geq 1, \tag{4}$$

where $\Psi = 1+\phi = \frac{3+\sqrt{5}}{2} \approx 2.61$ and ϕ is the "Golden Section".

In [21] the authors create the binary logistic regression model as to find the optimal vector $\beta = [\beta_0, \beta_1, \ldots, \beta_n]$ that best fits

$$y = \begin{cases} 1, & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon > 0 \\ 0, & \text{otherwise,} \end{cases}$$

here ε represents the error.

Evidently, in (1) t can be regarded as a variable, which is a linear weighted combination of independent variable $x = [x_1, \ldots, x_n]$ as

$$t \leftarrow \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n.$$

Thus, the binary logistic model is [21]:

$$F(x) = \frac{1}{1 + e^{-t(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}, \quad (5)$$

where $F(x)$ represents the probability of dependent variable $y = 1$.

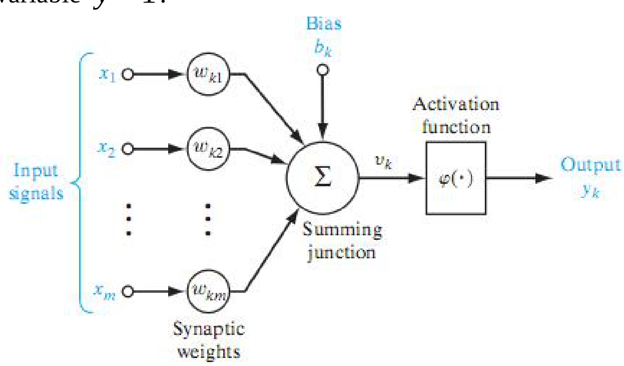

Fig. 1. Nonlinear, parametrized function with restricted output range [5].

Training a multilayer perceptron with algorithms employing global search strategies has been an important research direction in the field of neural networks.

Multi–layer perceptrons are feed forward neural networks featuring universal approximation properties used both in regression problems.

The standard feed forward networks with only a single hidden layer can approximate any continuous function uniformly on any compact set and any measurable function to any desired degree of accuracy [22]–[25].

The nonlinear, parametrized function with restricted output range is visualized on Fig.1.

It is straightforward to extend this analysis to networks with multiple hidden layers.

For recurrent neural networks are typical:
a) stable outputs may be more difficult to evaluate;
b) unexpected behavior (chaos, oscillation).

A survey of neural transfer activation functions can be found in [26].

Moreover, the nodes in the hidden layer are supposed to have a sigmoidal activation function, which may be one of the following:

a) logistic sigmoid

$$\varphi_1(net) = \frac{1}{1 + e^{-\beta net}}; \quad (6)$$

b) hyperbolic tangent

$$\varphi_2(net) = \frac{e^{\beta net} - e^{-\beta net}}{e^{\beta net} + e^{-\beta net}}; \quad (7)$$

c) half hyperbolic tangent

$$\varphi_3(net) = \frac{1 - e^{-\beta net}}{1 + e^{-\beta net}}; \quad (8)$$

d) Parametric Fibonacci hyperbolic tangent

$$\varphi_4(net) = \frac{\Psi^{\beta net} - \Psi^{-\beta net}}{\Psi^{\beta net} + \Psi^{-\beta net}}, \quad (9)$$

where net denotes the input to a node and β is the slope parameter of the sigmoids.

Definition 1 *The sign function of a real number* t *is defined as follows:*

$$sgn(t) = \begin{cases} -1, & \text{if} \quad t < 0, \\ 0, & \text{if} \quad t = 0, \\ 1, & \text{if} \quad t > 0. \end{cases} \quad (10)$$

Definition 2 *[27], [28] The Hausdorff distance (the H–distance) [27]* $\rho(f, g)$ *between two interval functions* f, g *on* $\Omega \subseteq \mathbb{R}$, *is the distance between their completed graphs* $F(f)$ *and* $F(g)$ *considered as closed subsets of* $\Omega \times \mathbb{R}$. *More precisely,*

$$\rho(f,g) = \max\{ \sup_{A \in F(f)} \inf_{B \in F(g)} \| A - B \|, \quad \sup_{B \in F(g)} \inf_{A \in F(f)} \| A - B \| \}, \quad (11)$$

wherein $\|.\|$ is any norm in $\mathbb{R}^2$, e. g. the maximum norm $\|(t, x)\| = \max\{|t|, |x|\}$; hence the distance between the points $A = (t_A, x_A)$, $B = (t_B, x_B)$ in $\mathbb{R}^2$ is $\| A - B \| = max(|t_A - t_B|, |x_A - x_B|)$.

In [29]–[34], [38] the authors consider some families of recurrence generated parametric activation functions on the base of (6)–(9).

A survey of new mathematical models of Nature is presented based on the Golden Section and using a class of hyperbolic Fibonacci and Lucas functions in [6].

3. Main Results

Using the $m(t;a,b,c,d)$ function to study some features in the antenna-feeder technic, the proportions between a,b,c and d are assumed.

It is natural to define the following modified function:

e) New Parametric Soboleva' modified hyperbolic tangent activation function (NPSMHTAF)

$$\varphi_5(t) = m(t;c,d,c,d) = \frac{e^{ct} - e^{-dt}}{e^{ct} + e^{-dt}} \quad (12)$$

or

$$\varphi_5(net) = \frac{e^{c\,net} - e^{-d\,net}}{e^{c\,net} + e^{-d\,net}}, \quad (13)$$

where net denotes the input to a node and c,d are the slope parameters of the sigmoid.

In this Section we prove upper and lower estimates for the Hausdorff approximation of the sign function by means of $\varphi_5(t)$.

3.1 Approximation issues

The H-distance $d_1(sgn(t),\varphi_5(t))$ between the sgn function and the function φ_5 satisfies the relation:

$$\varphi_5(d_1) = \frac{e^{cd_1} - e^{-dd_1}}{e^{cd_1} + e^{-dd_1}} = 1 - d_1. \quad (14)$$

The following Theorem gives upper and lower bounds for d_1

Theorem 3.1. *For the Hausdorff distance* d_1 *between the sgn function and the function* φ_5 *the following inequalities hold for* $c+d > 12.0189$,

$$d_{l_1} = \frac{1}{2\left(1+\frac{c+d}{2}\right)} <$$
$$d_1 < \frac{\ln\left(2\left(1+\frac{c+d}{2}\right)\right)}{2\left(1+\frac{c+d}{2}\right)} = d_{r_1}. \quad (15)$$

Proof. We define the functions

$$F(d_1) = \frac{e^{cd_1} - e^{-dd_1}}{e^{cd_1} + e^{-dd_1}} - 1 + d_1, \quad (16)$$

$$G(d_1) = -1 + \left(1+\frac{c+d}{2}\right)d_1. \quad (17)$$

From Taylor expansion we find

$$F(d_1) - G(d_1) = O(d_1^2).$$

In addition $G'(d_1) > 0$ and for $c+d > 12.0189$:

$$G(d_{l_1}) < 0;\ G(d_{r_1}) > 0.$$

This completes the proof of the inequalities (15).

Approximations of the $sgn(t)$ by (NPSMHTAF)-functions for various c and d are visualized on Fig. 2–Fig. 4

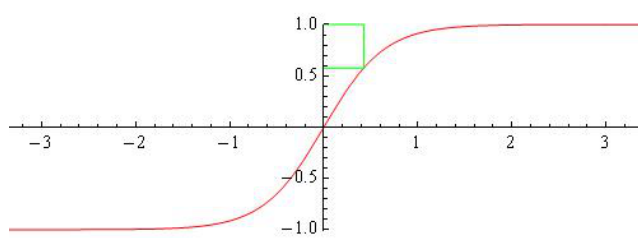

Fig. 2. Approximation of the $sgn(t)$ by (NPSMHTAF) for $c = 1.5$, $d = 1.6$; Hausdorff distance: $d_1 = 0.423756$.

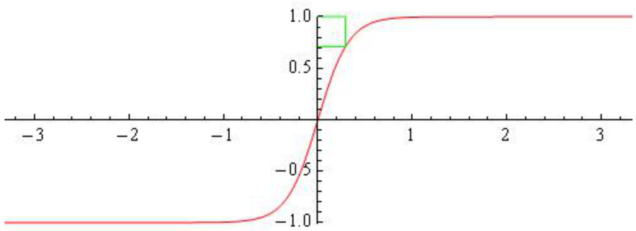

Fig. 3. Approximation of the $sgn(t)$ by (NPSMHTAF) for $c = 2.5$, $d = 3.6$; Hausdorff distance: $d_1 = 0.29053$.

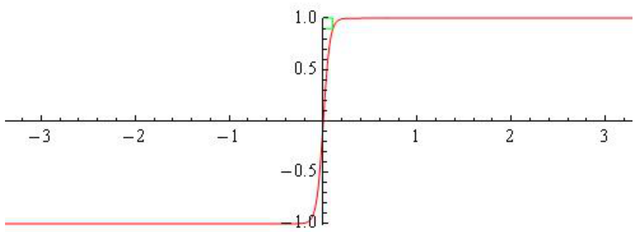

Fig. 4. Approximation of the $sgn(t)$ by (NPSMHTAF) for $c = 3.5$, $d = 25$; Hausdorff distance: $d_1 = 0.102427$.

Some computational examples using relations (12) are presented in Table 1.

The last column of Table 1 contains the values of d_1 computed by solving the nonlinear equation (14).

Table 1: Bounds for d_1 computed by (15) for various c and d.

c	d	d_{l_1}	d_{r_1}	d_1 from (14)
6	15	0.0434783	0.136326	0.127821
10	30	0.0238095	0.0889921	0.0795863
2	100	0.00961538	0.0446576	0.03852961
3.5	25	0.0327869	0.112057	0.102427

From the above table, it can be seen that the right estimates for the value of the best Hausdorff distance (see formula (15)) are quite precise.

4. The Family of Recurrence Generated Parametric Soboleva' Modified Hyperbolic Tangent Activation Function (FPSMHTAF)

We consider the following family of recurrence generated (FPSMHTAF) functions:

$$\delta_{i+1}(t) = \frac{e^{c(t+\delta_i(t))} - e^{-d(t+\delta_i(t))}}{e^{c(t+\delta_i(t))} + e^{-d(t+\delta_i(t))}}, \quad i = 0,1,2,\ldots, \qquad (18)$$

with

$$\delta_0(t) = \frac{e^{ct} - e^{-dt}}{e^{ct} + e^{-dt}}; \quad \delta_0(0) = 0. \qquad (19)$$

Evidently, $\delta_{i+1}(0) = 0$ for $i = 0,1,2,\ldots,$.

Denote the number of recurrences by p.

The recurrence generated (FPSMHTAF)–functions: $\delta_0(t)$, $\delta_1(t)$ and $\delta_2(t)$ for fixed $c = 3$, $d = 2.8$ and $p = 2$ are visualized on Fig. 5.

From the graphics it can be seen that the "saturation" is faster.

Remark. Based on the methodology proposed in the present note, the reader may formulate the corresponding approximation problems on his/her own.

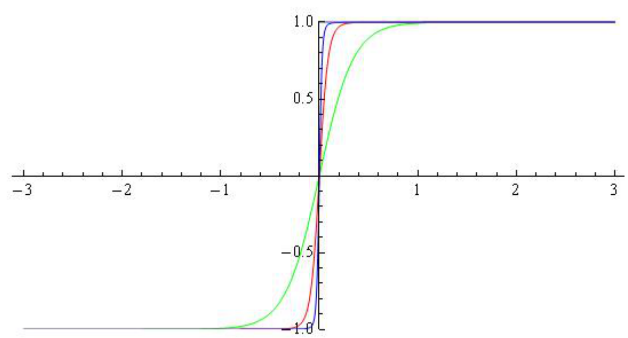

Fig. 5. Approximation of the $sgn(t)$ by (FPSMHTAF) for fixed $c = 3$, $d = 2.8$ and $p = 2$; The graphics of recurrence generated functions: δ_0 (green), δ_1 (red) and δ_2 (blue).

5. Conclusions

A family of recurrence generated parametric Soboleva' modified hyperbolic tangent activation function (FPSMHTAF) is introduced finding application in neural network theory and practice.

Theoretical and numerical results on the approximation in Hausdorff sense of the sgn function by means of functions belonging to the family are reported in the paper.

We propose a software module within the programming environment *CAS Mathematica* for the analysis of the considered family of recurrence generated (FPSMHTAF) functions.

The module offers the following possibilities:

- generation of the activation functions under user defined values of the parameters c, d and number of recursions p;

- calculation of the H-distance d_p, $p = 0,1,2,\ldots,$ between the sgn function and the activation functions $\delta_0, \delta_1, \delta_2, \ldots, \delta_p$;

- software tools for animation and visualization.

For other results, see [35]–[37].

Acknowledgments

This work has been supported by the project FP17-FMI-008 of Department for Scientific Research, Paisii Hilendarski University of Plovdiv.

References

[1] E. V. Soboleva, V. V. Beskorovainyi, The utility function in problems of structural optimization of distributed objects, Kharkiv University of Aair Force, 121, 2008.
[2] E. V. Soboleva, The S-shaped utility function of individual criteria for multi-objective decision–making in design, Kharkiv National University of Radioelectroniks, 247, 2009.
[3] E. V. Soboleva, V. V. Beskorovainyi, Identification of utility functions in multi–objective choice modelling by using S–shaped functions (PDF), Kharkiv National University of Radioelectroniks, 2010, 50–54.
[4] V. I. Tuev, M. V. Uzhanin, Using modified hyperbolic tangent function to approximate the current–voltage characteristics of field–effect transistors, Tomsk Polytechnic University, 2009, 135–138.
[5] S. Haykin, Neural networks and learning machines, – 3rd ed., Copyright by Pearson Education, Inc., Upper Saddle River, New Jersey 07458, 2009.
[6] A. Stakhov, I. Tkachenko, Hyperbolic Fibonacci Trigonometry, Dokl. Akad. Nauk Ukrainy, 7, 1993, 9–14.
[7] Z. Trzaska, On Fibonacci hyperbolic trigonometry and modified numerical triangles, Fib. Quart., 34, 1996, 129–138.
[8] N. Guliyev, V. Ismailov, A single hidden layer feedforward network with only one neuron in the hidden layer san approximate any univariate function, Neural Computation, 28, 2016, 1289–1304.
[9] D. Costarelli, R. Spigler, Approximation results for neural network operators activated by sigmoidal functions, Neural Networks, 44, 2013, 101–106.
[10] D. Costarelli, G. Vinti, Pointwise and uniform approximation by multivariate neural network operators of the max-product type, Neural Networks, 2016, doi:10.1016/j.neunet.2016.06.002.
[11] D. Costarelli, R. Spigler, Solving numerically nonlinear systems of balance laws by multivariate sigmoidal functions approximation, Computational and Applied Mathematics 2016, doi:10.1007/s40314–016–0334–8.
[12] D. Costarelli, G. Vinti, Convergence for a family of neural network operators in Orlicz spaces, Mathematische Nachrichten 2016, doi:10.1002/mana.20160006
[13] J. Dombi, Z. Gera, The Approximation of Piecewise Linear Membership Functions and Lukasiewicz Operators, Fuzzy Sets and Systems, 154(2), 2005, 275–286.
[14] I. A. Basheer, M. Hajmeer, Artificial Neural Networks: Fundamentals, Computing, Design, and Application, Journal of Microbiological Methods, 2000, 43, 3–31, doi:10.1016/S0167–7012(00)00201-3.
[15] Z. Chen, F. Cao, The Approximation Operators with Sigmoidal Functions, Computers & Mathematics with Applications, 58, 2009, 758–765, doi:10.1016/j.camwa.2009.05.001.
[16] Z. Chen, F. Cao, The Construction and Approximation of a Class of Neural Networks Operators with Ramp Functions, Journal of Computational Analysis and Applications, 14, 2012, 101–112.
[17] Z. Chen, F. Cao, J. Hu, Approximation by Network Operators with Logistic Activation Functions, Applied Mathematics and Computation 256, 2015, 565–571, doi:10.1016/j.amc.2015.01.049.
[18] D. Costarelli, R. Spigler, Constructive Approximation by Superposition of Sigmoidal Functions, Anal. Theory Appl. 29, 2013, 169–196, doi:10.4208/ata.2013.v29.n2.8.
[19] D. Elliott, A better activation function for artificial neural networks, the National Science Foundation, Institute for Systems Research, Washington, DC, ISR Technical Rep. TR–6, 1993.
[20] K. Babu, D. Edla, New algebraic activation function for multi-layered feed forward neural networks. IETE Journal of Research, 2016, doi:10.1080/03772063.2016.1240633.
[21] S. Wang, T. Zhan, Y. Chen, Y. Zhang, M. Yang, H. Lu, H. Wang, B. Liu, P. Phillips, Multiple Sclerosis Detection Based on Biorthogonal Wavelet Transform, RBF Kernel Principal Component Analysis, and Logistic Regression,

IEEE Access, Special section on advanced signal processing methods in medical imaging, 4, 2016, 7567–7576.

[22] G. Cybenko, Approximation by superposition of a sigmoidal function, Math. of Control Signals and Systems, 2, 1989, 303–314.

[23] K. Hornik, M. Stinchcombe, H. White, Multi-layer feed forward networks are universal approximations, Neural Networks, 2, 1989, 359–366.

[24] V. Kreinovich, O. Sirisaengtaksin, 3–layer neural networks are universal approximations for functionals and for control strategies, Neural Parallel and Scientific Computations, 1, 1993, 325–346.

[25] H. White, Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings, Neural Networks, 3, 1990, 535–549.

[26] W. Duch, N. Jankowski, Survey of neural transfer functions, Neural Computing Surveys, 2, 1999, 163–212.

[27] F. Hausdorff, Set Theory (2 ed.) (Chelsea Publ., New York, (1962 [1957]) (Republished by AMS-Chelsea 2005), ISBN: 978–0–821–83835–8.

[28] B. Sendov, Hausdorff Approximations, Kluwer, Boston, 1990, doi:10.1007/978–94–009–0673–0.

[29] N. Kyurkchiev, A family of recurrence generated sigmoidal functions based on the Verhulst logistic function. Some approximation and modeling aspects, Biomath Communications, 3(2), 2016, 18 pp.

[30] A. Iliev, N. Kyurkchiev, S. Markov, A family of recurrence generated parametric activation functions with applications to neural networks, International Journal on Research Innovations in Engineering Science and Technology, 2(1), 2017, 60–68.

[31] N. Kyurkchiev, S. Markov, Hausdorff Approximation of the Sign Function by a Class of Parametric Activation Functions, Biomath Communications, 3(2), 2016, 14 pp., doi:10.11145/bmc.2016.12.217.

[32] N. Kyurkchiev, A. Iliev, S. Markov, Families of recurrence generated three and four parametric activation functions, Int. J. Sci. Res. and Development, 4(12), 2017, 746–750.

[33] V. Kyurkchiev, N. Kyurkchiev, Afamily of recurrence generated functions based oh Half-hyperbolic tangent activation functions, Biomedical Statistics and Informatics, 2(3), 2017, 87–94.

[34] V. Kyurkchiev, A. Iliev, N. Kyurkchiev, On some families of recurrence generated activation functions, Int. J. of Sci. Eng. And Appl. Sci., 3(3), 2017.

[35] N. Kyurkchiev, S. Markov, Sigmoid functions: Some Approximation and Modelling Aspects, LAP LAMBERT Academic Publishing, Saarbrucken, 2015, ISBN 978–3–659–76045–7.

[36] N. Kyurkchiev, S. Markov, On the Hausdorff distance between the Heaviside step function and Verhulst logistic function. J. Math. Chem., 54(1), 2016, 109–119, doi:10.1007/S10910–015–0552–0.

[37] N. Kyurkchiev, A. Iliev, On the Hausdorff distance between the Heaviside function and some transmuted activation functions, Mathematical Modelling and Applications, 2(1), 2016, 1–5.

[38] N. Kyurkchiev, A. Iliev, A note on the new Fibonacci hyperbolic tangent activation function, Int. J. of Innovative Science, Engineering and technology, 4(5), 2017, 364–368.