Comments and Replies

Comments on "The Extreme Learning Machine"

Lipo P. Wang and Chunru R. Wan

Abstract—**This comment letter points out that the essence of the "extreme learning machine (ELM)" recently appeared has been proposed earlier by Broomhead and Lowe and Pao *et al.*, and discussed by other authors. Hence, it is not necessary to introduce a new name "ELM."**

Index Terms—**Fast training, fixed centers, pseudoinverse, random neurons, random vector functional link.**

Recently, Huang *et al.* proposed an "extreme learning machine (ELM)" (see [1] and its references 7–10, 12, and 13; and [2]), which is a single-hidden-layer feedforward neural network with the hidden neuron parameters [i.e., weights and bias for a multilayer perceptron (MLP) network, centers and width's in a radial basis function (RBF) network] randomly assigned and only the output weights adjusted during training. The output weights can be adjusted in one of the following ways [1]: 1) using pseudoinverse (also known as Moore–Penrose generalized inverse); 2) incrementally (at each iteration, a new random hidden neuron is added); or 3) online sequentially (as new data arrive in real-time applications) [2]. Huang *et al.* [1] also proved that an incremental "ELM" is a universal approximator.

The idea that hidden RBF neurons randomly selected from the domain of a data space are sufficient to allow universal approximation was proposed by Broomhead and Lowe in their classic paper [3] (with more than 580 citations in the ISI Web of Science) in 1988 and later extended by Lowe [4] (also see textbooks by Haykin [5] and Ham and Kostanic [6]). Broomhead and Lowe demonstrated that unless one is interested in a minimal network with the least number of RBF neurons, one can simply fix the hidden RBF neurons and only train the output weights, which reduces network training to a nearly trivial problem of linear least square minimization. The output weights can then be determined by pseudoinverse, which is fast and efficient. They considered RBF centers selected uniformly over a grid in the domain, and introduced the idea of randomizing this process and using fewer centers than data points. RBF centers may or may not be randomly selected data points, as stated in the footnote [3, p. 325], "we do not necessarily require that the radial basis function centers correspond to any of the data points." Lowe [4] further said that it is sensible to select the RBF centers from the training data (it would not make much practical sense to select RBFs outside the domain of the training data, which is a matter of data scale). This RBF network with randomly fixed hidden neurons (RHN) has been used by other authors. For example, in solving the two-spiral benchmark, centers of the basis functions were drawn randomly from a 2-D Gaussian distribution centered at the spirals foci ([7, p. 1181]). Wettschereck and Dietterich [8] compared the performance of the RHN RBF network, the *general* RBF network (with fully adjustable hidden neuron parameters and output weights), and the *general* MLP (with fully adjustable hidden and output neuron weights and biases).

As a special case of their more general functional-link neural networks (FLN) described in Pao's popular textbook [9] (cited over 700 times in the ISI Web of Science), Pao *et al.* (e.g., [10]–[15]) proposed the random vector functional-link (RVFL) network where the hidden neurons in an FLN are randomly selected and only the weights of the output layer need to be trained (e.g., with pseudoinverse or gradient descent), with the only difference from the "ELM" being that the RVFL allows for direct connections from the input nodes to the output neurons, whereas the "ELM" does not. Igelnik and Pao [12] proved that the RVFL network is a universal approximator. Igelnik, Pao, LeClair, and Shen [14] discussed learning and generalization with a one-hidden-layer feedforward neural network consisting of heterogeneous and randomly prescribed nodes. Pao *et al.* showed that the RVFL is fast and accurate, as also established in various applications by other researchers (e.g., [16]–[34]; also see the textbook by Looney [35]). In particular, Lewis and co-workers [21]–[24] demonstrated that the RVFL makes efficient neural controllers. Husmeier [18] and Taylor [19], [20] used the RVFL with the expectation–maximization (EM) algorithm for probability–density estimation. Chen and Wan [27] proposed both incremental (with added neurons) and online sequential (with new training data) learning algorithms for this type of networks suited for real-time applications. MLP neurons in the RVFL have generalized to other types of neurons, such as RBF [28], [29] and trigonometric functions (e.g., sin and cos) neurons [30].

In conclusion, feedforward networks (both RBF and MLP) with randomly fixed hidden neurons (RHN) have previously been proposed and discussed by other authors in papers and textbooks. These RHN networks have been shown, both theoretically and experimentally, to be fast and accurate. Hence, it is not necessary to introduce a new name "ELM."

Manuscript received October 2, 2006; revised February 23, 2007 and February 1, 2008; accepted May 4, 2008. Published August 6, 2008 (projected).

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: elpwang@ntu.edu.sg, ecrwan@ntu.edu.sg).

Digital Object Identifier 10.1109/TNN.2008.2002273

References

[1] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.

[2] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1411–1423, Nov. 2006.

[3] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Syst.*, no. 2, pp. 321–355, 1988.

[4] D. Lowe, "Adaptive radial basis function nonlinearities, and the problem of generalisation," in *Proc. 1st Inst. Electr. Eng. Int. Conf. Artif. Neural Netw.*, 1989, pp. 171–175.

[5] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[6] F. M. Ham and I. Kostanic, *Principles of Neurocomputing for Science & Engineering*. Boston, MA: McGraw-Hill, 2001.

[7] W. Kaminski and P. Strumillo, "Kernel orthonormalization in radial basis function neural networks," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 1177–1183, Sep. 1997.

[8] D. Wettschereck and T. Dietterich, "Improving the performance of radial basis function networks by learning center locations," in *Advances in Neural Information Processing Systems (NIPS)*. San Mateo, CA: Morgan Kaufmann, 1992, vol. 4, pp. 1133–1140.

[9] Y. H. Pao, *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley, 1989.
[10] Y. H. Pao and Y. Takefuji, "Functional-link net computing, theory, system architecture, and functionalities," *IEEE Comput.*, vol. 25, no. 5, pp. 76–79, May 1992.
[11] Y. H. Pao, G. H. Park, and D. J. Sobajic, "Learning and generalization characteristics of random vector functional-link net," *Neurocomputing*, vol. 6, pp. 163–180, 1994.
[12] B. Igelnik and Y. H. Pao, "Stochastic choice of basis functions in adaptive function approximation and the functional-link net," *IEEE Trans. Neural Netw.*, vol. 6, no. 6, pp. 1320–1329, Nov. 1995.
[13] G.-H. Park, Y.-H. Pao, B. Igelnik, K. G. Eyink, and S. R. LeClair, "Neural-net computing for interpretation of semiconductor film optical ellipsometry parameters," *IEEE Trans. Neural Netw.*, vol. 7, no. 4, pp. 816–829, Jul. 1996.
[14] B. Igelnik, Y.-H. Pao, S. R. LeClair, and C. Y. Shen, "The ensemble approach to neural-network learning and generalization," *IEEE Trans. Neural Netw.*, vol. 10, no. 1, pp. 19–30, Jan. 1999.
[15] Z. Meng and Y. H. Pao, "Visualization and self-organization of multidimensional data through equalized orthogonal mapping," *IEEE Trans. Neural Netw.*, vol. 11, no. 4, pp. 1031–1038, Jul. 2000.
[16] H.-M. Chi and O. K. Ersoy, "A statistical self-organizing learning system for remote sensing classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1890–1900, Aug. 2005.
[17] G. H. Park, J. H. Park, and Y. J. Lee, "Intelligent rate control for MPEG-4 coders," in *Proc. 2nd Int. Conf. Intell. Process. Manuf. Mater.*, 1999, vol. 2, pp. 1285–1295.
[18] D. Husmeier, "The Bayesian evidence scheme for regularizing probability-density estimating neural networks," *Neural Comput.*, vol. 12, no. 11, pp. 2685–2717, Nov. 2000.
[19] D. Husmeier and J. G. Taylor, "Modelling conditional probabilities with committees of RVFL networks," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1997, vol. 1327, pp. 1053–1058.
[20] D. Husmeier and J. G. Taylor, "Neural networks for predicting conditional probability densities: Improved training scheme combining EM and RVFL," *Neural Netw.*, vol. 11, no. 1, pp. 89–116, Jan. 1998.
[21] R. R. Selmic and F. L. Lewis, "Deadzone compensation in motion control systems using neural networks," *IEEE Trans. Autom. Control*, vol. 45, no. 4, pp. 602–613, Apr. 2000.
[22] R. R. Selmic and F. L. Lewis, "Neural net backlash compensation with Hebbian tuning using dynamic inversion," *Automatica*, vol. 37, no. 8, pp. 1269–1277, Aug. 2001.
[23] O. Kuljaca, N. Swamy, F. L. Lewis, and C. M. Kwan, "Design and implementation of industrial neural network controller using backstepping," *IEEE Trans. Ind. Electron.*, vol. 50, no. 1, pp. 193–201, Feb. 2003.
[24] O. Kuljaca and F. L. Lewis, "Adaptive critic design using non-linear network structures," *Int. J. Adapt. Control Signal Process.*, vol. 17, no. 6, pp. 431–445, Aug. 2003.
[25] A. Lindemann, C. L. Dunis, and P. Lisboa, "Probability distributions, trading strategies and leverage: An application of Gaussian mixture models," *J. Forecasting*, vol. 23, no. 8, pp. 559–585, Dec. 2004.
[26] A. Lindemann, C. L. Dunis, and P. Lisboa, "Probability distributions and leveraged trading strategies: An application of Gaussian mixture models to the Morgan Stanley technology index tracking fund," *Quant. Finance*, vol. 5, no. 5, pp. 459–474, Oct. 2005.
[27] C. L. P. Chen and J. Z. Wan, "A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction," *IEEE Trans. Syst. Ma, Cybern. B, Cybern.*, vol. 29, no. 1, pp. 62–72, Feb. 1999.
[28] A. Hernández-Aguirre, C. Koutsougeras, and B. P. Buckles, "Sample complexity for function learning tasks through linear neural networks," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2002, vol. 2313, pp. 262–271.
[29] A. Porwal, E. J. M. Carranza, and M. Hale, "Artificial neural networks for mineral-potential mapping: A case study from Aravalli Province, Western India," *J. Natural Resources Res.*, vol. 12, no. 3, pp. 155–171, Sep. 2003.
[30] D. P. Das and G. Panda, "Active mitigation of nonlinear noise processes using a novel filtered-s LMS algorithm," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 3, pp. 313–322, May 2004.
[31] P. G. He and S. Jagannathan, "Neuro-controller for reducing cyclic variation in lean combustion spark ignition engines," *Automatica*, vol. 41, no. 7, pp. 1133–1142, Jul. 2005.
[32] P. G. He and S. Jagannathan, "Reinforcement learning-based output feedback control of nonlinear systems with input constraints," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 35, no. 1, pp. 150–154, Feb. 2005.
[33] W. Z. Gao and R. R. Selmic, "Neural network control of a class of nonlinear systems with actuator saturation," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 147–156, Jan. 2006.
[34] B. S. Lin, F.-C. Chong, and F. Lai, "A functional link network with higher order statistics for signal enhancement," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4821–4826, Dec. 2006.
[35] C. L. Looney, *Pattern Recognition Using Neural Networks*. Oxford, U.K.: Oxford Univ. Press, 1997.

Reply to "Comments on "The Extreme Learning Machine""

Guang-Bin Huang

In this reply, we refer to Wang and Wan's comments on our research publication. We found that the comment letter contains some inaccurate statements. The comment letter contains some contradictions as well.

The meaning of "random units" varies in different authors' works. We have comprehensively compared extreme learning machine (ELM), Lowe's work, random vector functional link (RVFL), and others in our work [2], [3].

The comment letter stated that "The idea that hidden RBF neurons randomly selected from the domain of a data space are sufficient to allow universal approximation was proposed by Broomhead and Lowe in their classic paper (1988)." This statement is incorrect because of the following.

1) Lowe's work does not randomly select the RBF neurons (i.e., all the parameters of hidden neurons as done in ELM). Instead, Lowe's work only randomly selects the radial basis function (RBF) centers but not the impact factors.
2) Lowe's work does not address the universal approximation but only focuses on the data interpolation. Interestingly, Lowe's work will not have the universal approximation capability if it moves one more step towards ELM direction. The *same* impact factor b is selected heuristically for all the hidden nodes in Lowe's network $f_n(\mathbf{x}) = \sum_{i=1}^{n} \beta_i g(b\|\mathbf{x} - \mathbf{a}_i\|)$. If one randomly chooses b, then the universal approximation capability of such a network is lost. The authors of the comment letter also have missed the apparent difference between Lowe's network $f_n(\mathbf{x}) = \sum_{i=1}^{n} \beta_i g(b\|\mathbf{x} - \mathbf{a}_i\|)$ and our ELM network $f_n(\mathbf{x}) = \sum_{i=1}^{n} \beta_i G(\mathbf{x}, \mathbf{a}_i, b_i)$.

This comment letter also stated that Haykin's textbook mentioned the universal approximation of Lowe's work. However, Haykin's textbook does not mention the universal approximation of Lowe's work at all. Instead, it only shows the interpolation ability of Lowe's work. The difference between the universal approximation and interpolation is important indeed.

Manuscript received May 17, 2007; revised February 1, 2008; accepted May 4, 2008. Published August 6, 2008 (projected).

The author is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: gbhuang@ieee.org).

Digital Object Identifier 10.1109/TNN.2008.2002275