

CASCADE ERROR PROJECTION _A LEARNING ALGORITHM FOR HARDWARE IMPLEMENTATION

Tuan A. Duong and Taher Daud
Center for Space Microelectronics Technology
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, CA 91109

Abstract:

In this paper, we work out a detailed mathematical analysis for a new learning algorithm termed Cascade Error Projection (CEP) and a general learning frame work. This frame work can be used to obtain the cascade correlation learning algorithm by choosing a particular set of parameters. Furthermore, CEP learning algorithm is operated only on one layer, whereas the other set of weights can be calculated deterministically. In association with the dynamical stepsize change concept to convert the weight update from infinite space into a finite space, the relation between the current stepsize and the previous energy level is also given and the estimation procedure for optimal stepsize is used for validation of our proposed technique.

The weight values of zero are used for starting the learning for every layer, and a single hidden unit is applied instead of using a pool of candidate hidden units similar to cascade correlation scheme. Therefore, simplicity in hardware implementation is also obtained. Furthermore, this analysis allows us to select from other methods (such as the conjugate gradient descent or the Newton's second order) one of which will be a good candidate for the learning technique. The choice of learning technique depends on the constraints of the problem (e.g., speed, performance, and hardware implementation); one technique may be more suitable than others. Moreover, for a discrete weight space, the theoretical analysis presents the capability of learning with limited weight quantization. Finally, 5- to 8-bit parity and chaotic time series prediction problems are investigated; the simulation results demonstrate that 4-bit or more weight quantization is sufficient for learning neural network using CEP. In addition, it is demonstrated that this technique is able to compensate for less bit weight resolution by incorporating additional hidden units. However, generation result may suffer somewhat with lower bit weight quantization.

I-Introduction

There are many ill-defined problems in pattern recognition, classification, vision, and speech recognition which need to be solved in real time [1-3]. One of the most attractive features of the neural network is a massively parallel processing topology that offers tremendous speed specially when implemented in hardware. Generally, neural network approaches in hardware face two main obstacles:

- (1) difficulty of network convergence due to the learning algorithm itself as well as the limited precision of the devices;
- (2) high cost of implementing hardware to truly mimic the synapse and neuron transfer functions dictated by the algorithm.

Furthermore, the convergence and the implementable hardware have a mutual correlation to each other; for example, the convergence of the learning network depends on the weight resolution available in synapse [4-6], and the cost of implementation of each bit in synapse grows, at least doubly, in silicon area, power, and connectivity [7-8]

In this paper, CEP learning algorithm is presented. It offers a simple learning method using a one-layer perceptron approach and a deterministic calculation for the other layer. Such a simple procedure offers a fast, reliable, and implementable learning algorithm. In addition, the learning technique is not only tolerant of 3- and 4-bit weight

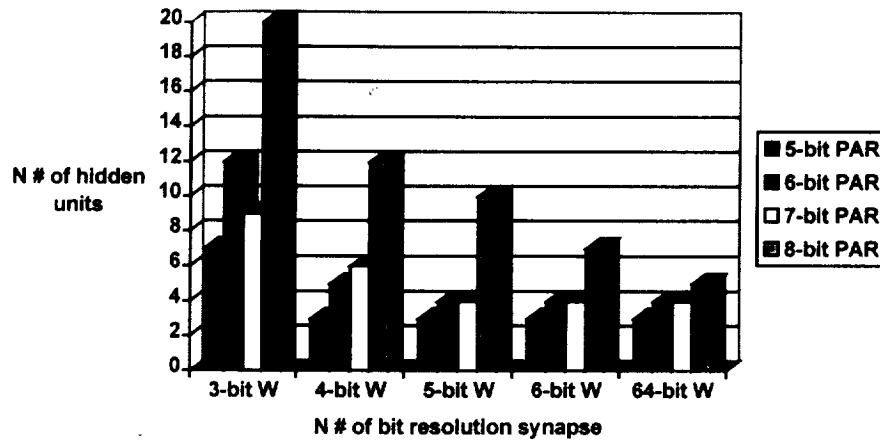


Figure 2: The chart shows CEP learning capability and the number of hidden units required to correctly solve 5- to 8-bit parity problems using round-off technique. x axis represents weight quantization (3-6 and 64-bit) and y axis shows the resulting number of hidden units (limited to 20). Each learning hidden unit is provided with 100 epoch iterations. As shown, a larger number of hidden units compensate for the lower weight resolution.

Chaotic Time Series Problem:

The data in this problem represents chaos and never repeated. However, this data between past, present, and future are correlated in high order. To validate the capability of CEP as shown in theory, we use CEP learning technique under constraints of limited weight quantization (4-, 6-, and 64-bit weight resolution) to capture the high order correlation of this problem.

In this experiment, we use $x_i, x_{i+1}, x_{i+2}, x_{i+3}$ and the target is x_{i+4} . The number of training data is 351 and test data is 651 and no cross validating data is applied in this phase.

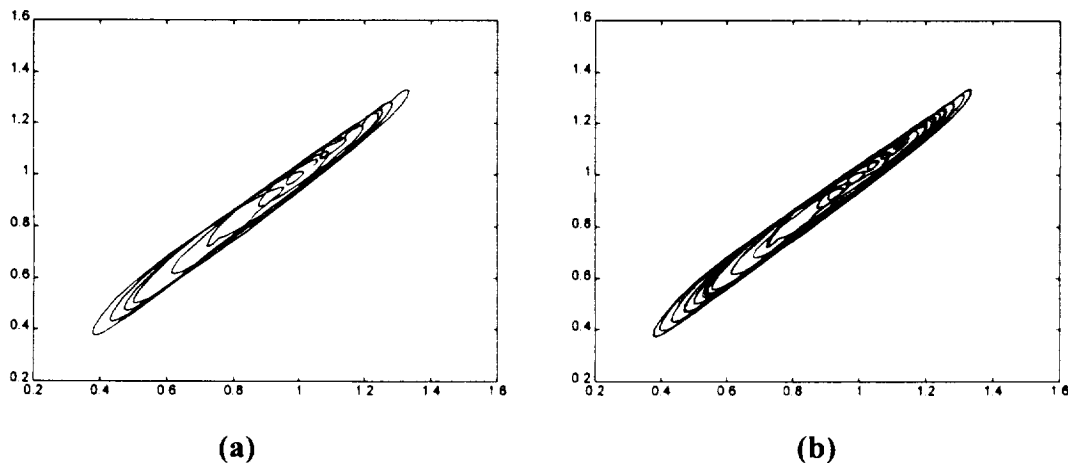


Figure 3: Data sets of chaotic time series problem. (a). training set to the CEP neural network, and (b). Test set which has no overlap with training set.

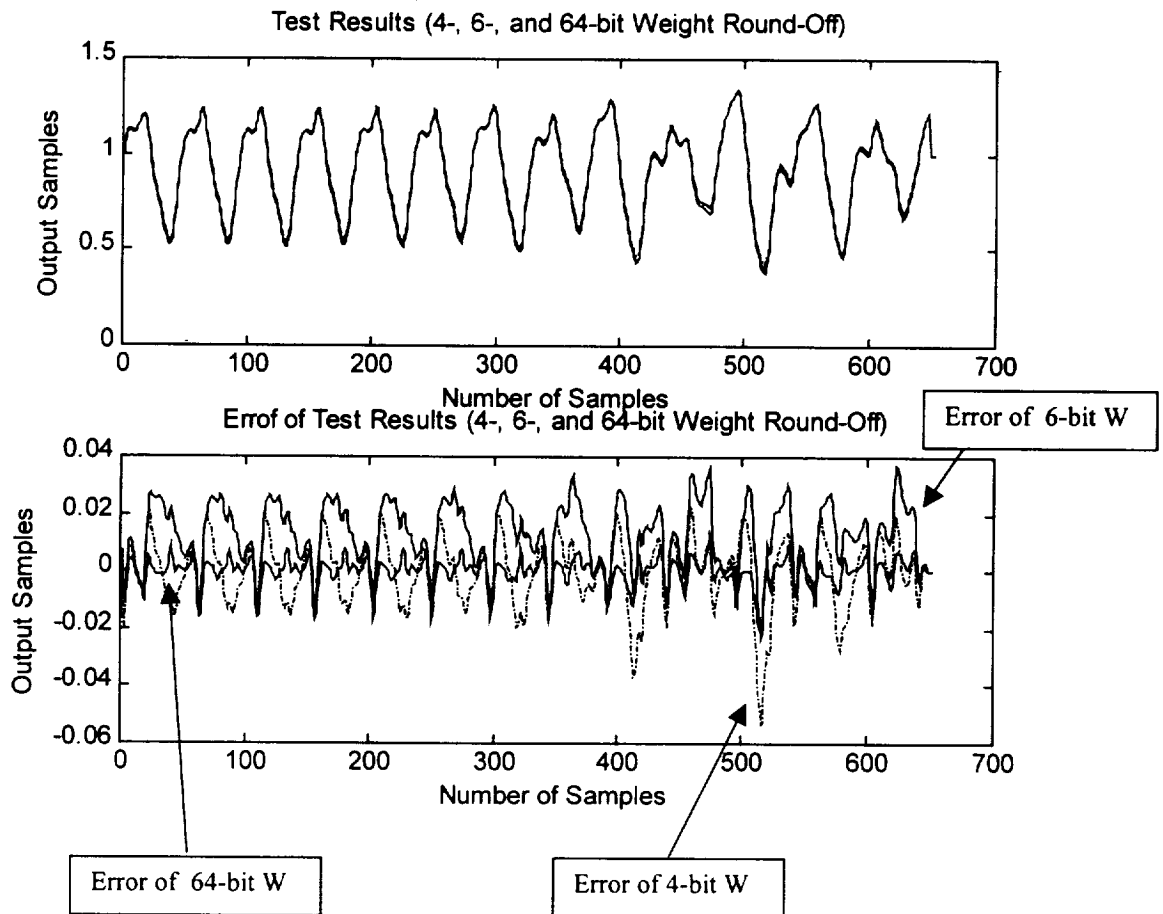


Figure 4: Simulation Results of CEP for chaotic time series prediction problem. Top trace contains four curves: ideal data, 64-bit, 6-bit and 4-bit prediction results. Bottom trace contains : errors between ideal data and 64-bit, 6-bit, and 4-bit generalization data.

The results in Figure 4 show that the error between ideal data and prediction with 64-bit weight learning network is within ± 0.01 and is like white noise, whereas, 6-bit error is more harmonic than 4-bit error prediction. These results can be interpreted to infer that the more bit weight quantization is available for learning the better and smoother the transform would be. In addition, the better and smoother transformation will help network to interpolate for predictions.

IV. Conclusions

In this paper, we have shown that CEP is a reliable technique for both software- and hardware-based neural network learning. From this analysis, it is shown that the CC algorithm is a special case and can be understood in greater depth with this analysis. Moreover, the theoretical analysis provides us with the general framework of the learning

architecture, and the particular learning algorithm can be independently studied for its suitability for a given application associated with given constraints specific to each problem. For example, for hardware implementation CEP is advantageous, but for software, covariance or Newton's second order method is more advantageous). For the CEP learning algorithm, the advantages can be summarized as follows:

- A fast and reliable learning technique
- A hardware implementable learning technique
- Learning scheme is tolerant of lower weight resolutions.
- A robust model in learning neural networks

Acknowledgments:

The research described herein was performed by the Center for Space Microelectronics Technology, Jet Propulsion Laboratory, California Institute of Technology and was jointly sponsored by the Ballistic Missile Defense Organization/Innovative Science and Technology Office (BMDO/IST), and the National Aeronautics and Space Administration (NASA). The authors would like to thank Drs A. Stubberud and A. Thakoor for useful discussions.

References:

- [1] T. A. Duong, T. Brown, M. Tran, H. Langenbacher, and T. Daud, "Analog VLSI neural network building block chips for hardware-in-the-loop learning," *Proc. IEEE/INNS Int'l Join Conf. on Neural Networks*, Beijing, China, Nov. 3-6, 1992.
- [2] T. A. Duong et. al, "Low Power Analog Neurosynapse Chips for a 3-D "Sugarcube" Neuroprocessor," *Proc. of IEEE Intl' Conf. on Neural Networks(ICNN/WCCI)*, Vol III, pp. 1907-1911, June 28-July 2, 1994, Orlando, Florida.
- [3] B.E. Boser, E. Sackinger, J. Bromley, Y. LeCun, and L.D. Jackel, "An Analog Neural Network Processor with Programmable Topology," *IEEE Journal of Solid State Circuits*, vol. 26, NO. 12, Dec. 1991.
- [4] P. W. Hollis, J.S. Harper, and J.J. Paulos, "The effects of Precision Constraints in a Backpropagation learning Network," *Neural Computation*, vol. 2, pp. 363-373, 1990.
- [5] M. Hoehfeld and S. Fahlman, "Learning with limited numerical precision using the cascade-correlation algorithm," *IEEE Trans. Neural Networks*, vol.3, No. 4, pp 602-611, July 1992.
- [6] T.A. Duong, S.P. Eberhardt, T. Daud, and A. Thakoor, "Learning in neural networks: VLSI implementation strategies," In: *Fuzzy logic and Neural Network Handbook*, Chap. 27, Ed: C.H. Chen, McGraw-Hill, 1996.
- [7] S.P. Eberhardt, T.A. Duong, and A.P. Thakoor, "Design of parallel hardware neural network systems from custom analog VLSI "building-block" chips," *IEEE/INNS Proc. IJCNN*, June 18-22, 1989 Washington D.C., vol. II, pp. 183.
- [8] T. A. Duong, S. P. Eberhardt, M. D. Tran, T. Daud, and A. P. Thakoor, "Learning and Optimization with Cascaded VLSI Neural network Building-Block Chips," *Proc. IEEE/INNS International Join Conference on Neural Networks*, June 7-11, 1992, Baltimore, MD, vol. I, pp. 184-189.
- [9] T. A. Duong, *Cascade Error Projection_ An sufficient Hardware learning theory. Ph.D. Thesis, UCI, 1995.*

- [10] S. E. Fahlmann, C. Lebiere, "The Cascade Correlation learning architecture," in *Advances in Neural Information Processing Systems II*, Ed: D. Touretzky, Morgan Kaufmann, San Mateo, CA, 1990, pp. 524-532.
- [11] T.A. Duong, "Cascade Error Projection-An efficient hardware learning algorithm," Proceeding Int'l IEEE/ICNN in Perth, Western Australia, vol. 1, pp. 175-178, Oct. 27-Dec 1, 1995 (*Invited Paper*).
- [12] T.A. Duong, A. Stubberud, T. Daud, and A. Thakoor, "Cascade Error Projection-A New Learning Algorithm," Proceeding Int'l IEEE/ICNN in Washington D.C., vol. 1, pp. 229-234, Jun. 3-Jun 7, 1996.