International Journal of Information Processing, 6(3), 1-12, 2012
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

Measuring the Learnability of a Parameter in Multi-Layer Perceptron Neural Network

Siddhaling Urolagin[a], Prema K V[b] and N V Subba Reddy[c]

[a]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal, India, Contact: siddesh_u.@yahoo.com

[b]Mody Institute of Technology and Science, Rajasthan, India, Contact: drprema.mits@gmail.com

[c]Mody Institute of Technology and Science, Rajasthan, India, Contact: dr_nvsreddy@rediffmail.com

During the training process of Multi-Layer Perceptron (MLP) not all parameters will learn in the equal proportion; some are more discriminative than others. How network parameters are learning during the training process? The objective analysis of this question has many applications in the neural network, such as network freezing, network pruning, fault tolerance etc.. A method is developed in this research to measure the learning capability of a parameter called as learnability. The learnability is defined as the contribution of that parameter to reduce the network error as it undergoes training. The association between surface of performance index and learnability is studied. On the steeper regions, the learnability is higher than on the flatter regions of the surface. Further studies are carried out by removing a parameter based on learnability from a trained network. On removal of a parameter with higher learnability has drastically affected the Mean Square Error (MSE) and classification rate. On the other hand, a parameter with a lower learnability has least impact on MSE and classification rate of a trained network. In this research, the learnability measurement is applied on network freezing and network pruning to improve the performance of MLP network.

Keywords : Freezing, Learnability, Multi-Layer Perceptron, Pruning.

1. INTRODUCTION

Artificial Neural Network (ANN) is a biologically inspired model. Among the ANNs the backpropagation neural network (also called a Multi-Layer Perceptron (MLP)) is the most widely used network. There are many applications of MLP in the area of pattern recognition, computer vision, speech recognition and speech synthesis [1], *etc.*. To name a few, handwriting recognition [2][3], speech recognition [4][5], product inspection[6], optimization of computation [7], forecasting [8], fault detection [9], medical diagnosis [10][11] *etc.*. The MLP consist of a number of highly interconnected computing units, called neurons or nodes. Each unit receives inputs from other units in the network, or from the outside world and calculates an output based on these inputs. Associated with each connection (sometimes called a synapse) between the units is a weight. Each neuron has a bias associated with it. Typically, an architecture is chosen and the network undergoes training process to reduce the network error or performance index. As training progresses from initial weight and bias values, the network parameters such as weights and biases are updated according to the learning algorithm to reduce the performance index [12]. Not all the network parameters equally learn the input-output mapping. Some parameters would hold more discriminating capability while others are not so effective. It is interesting to investigate how the network parameters are the input-output mapping, some hold higher discrimination than others. In this research a method is developed to measure the learning capability (called as learnability) of a network parameter. Investigations are carried-out on the association of learnability and performance index's surface. It is found that learnability factors of parameters on the steeper region are higher than on the flatter region. Further, the behavior of learnability of several parameters during the training process is analyzed on MNIST handwritten database. In addition, the effect of removing a parameter based on learnability from a trained network is studied. It is observed that, the parameter having higher learnability has the more impact on MSE and classification rate of a network. In this research, learnability measurement has been applied on two important problems: network freezing and network pruning. The network freezing is one of the methods to improve the training time. It involves identifying the parameters those learnability is low and freezing of these parameters from the training. Once parameters are frozen, they will not participate in the training process. The network pruning is the method to improve the generalization performance by eliminating least significant parameters based on the learnability. Several experiments are conducted on feedforward neural network trained with backpropagation learning algorithm on MNIST handwritten numeral database. The network freezing and pruning results are compared with standard feedforward neural network where no freezing or no pruning is performed. The experimental results show better performance in terms of reduction in training times and improvement in classification rates and mean squared errors.

REFERENCES

1. G P Zhang. Neural Networks for Classification: A Survey, *in IEEE Transactions on Systems, Man, and Cybernetics*, part C, 30(4):451-462, 2000.
2. A. Krzyzak, W Dai, C Y Suen. Classification of Large Set of Handwritten Characters using Modified Backpropagation Model, *in International Joint Conference on Neural Networks*, 3:225-232, 1990.
3. Xin Wang, T Huang, X Liu. Handwritten Character Recognition Based on BP Neural Network, *in International Conference on Genetic and Evolutionary Computing*, pages 520-524, 2009.
4. S M Peeling, R K Moore and M J Tomlinson. The Multi-Layer Perceptron as a Tool for Speech Pattern Processing Research, *in Proceedings of IoA Autumn Conference on Speech and Hearing*, 8:307-314, 1986.
5. N Satravaha, P Klinkhachorn, N Lass. Tone Classification of Syllable-Segmented Thai Speech based on Multilayer Perceptron, *in Proceedings of the 35th Southeastern Symposium on System Theory*, pages 392-396, 2003.
6. S Sardy, L Ibrahim, Y Yasuda. A Application of Vision System for the Identification and Defect Detection on Woven Fabrics by using Artificial Neural Network, *in Proceedings of 1993 International Joint Conference on Neural Networks, (IJCNN '93-Nagoya)*, 3:2141-2144, 1993.
7. J Hopfield and D Tank. Neural Computation of Decisions in Optimization Problems, *in Biological Cybernetics*, 52:141-152, 1985.
8. K L Ho, Y Y Hsu, and C C Yang. Short Term Load Forecasting using a Multilayer Neural Network with an Adaptive Learning Algorithm, *in IEEE Transactions on Power Systems*, 7(1):141-149, 1992.
9. C Kim, Y Lim, W Chung, T Kwon, J Hwang, Il-D Kim. A study on the Fault Identification of Underground Cable using Neural Networks, *in Proceedings of International Conference on Energy Management and Power Delivery*, 2:571-576, 1995.
10. S Karamchandani, U B Desai, S N Merchant, G D Jindal. Principal Component Analysis based Backpropagation Algorithm for Diagnosis of Peripheral Aarterial Occlusive Diseases, *in Canadian Conference on Electrical and Computer Engineering, (CCECE '09)*, pages 482-485, 2009.
11. I Middleton, R I Damper. Identification of Boundaries in MRI Medical Images using Artificial Neural Networks, *in IEEE Colloquium on Artificial Intelligence Methods for Biomedical Data Processing*, 6:1-6, 1996.
12. Pierre F Baldi, Kurt Homik. Learning in Linear Neural Networks: A Survey, *in IEEE Transactions on Neural Networks*, 6(4), July

1995.
13. S S Erdogan, Geok-See Ng, P K-H Chan. Measurement Criteria for Neural Network Pruning, *in Proceedings of IEEE conference on Digital Signal Processing Applications*, 1:83-89, 1996.
14. R Reed. Pruning Algorithms: A Survey, *in IEEE Transactions on Neural Networks*, 4(5):740-747, 1993.
15. M Islam, Shahjahan, K Murase. A New Weight Freezing Method for Reducing Training Time in Designing Artificial Neural Networks, *in IEEE International Conference on Systems, Man, and Cybernetics*, 1:341-346, 2001.
16. Ching-Tai Chiu, K Mehrotra, C K Mohan, S Ranka. Modifying Training Algorithms for Improved Fault Tolerance, *in IEEE International Conference on Computational Intelligence*, pages 644-650, 1994.
17. B E Segee, M J Carter. Fault Tolerance of Pruned Multilayer Networks, *in Proceeding of International Joint Conference on Neural Networks, (IJCNN-91)*, Seattle, 2:447-452, 1991.
18. LeCun Y. Connectionist Learning Models, PhD thesis, University P.et M. Curie, 1987.
19. Le Cun Y, Denker J S and Solla S A. Optimal Brain Damage, *in Advances in Neural Information Processing Systems*, 2:598- 605, Morgan Kaufmann, 1989.
20. Simon Haykin. Neural Networks: A Comprehensive Foundation, Pearson Education, Singapore, 1999.
21. Siddhaling Urolagin, Prema K V, N V Subba Reddy. Robust Object Recognition using Binarized Gabor Features Under Noise and Illumination Changes, *in International Journal of Information Processing*, 2(2):170-182, 2008.
22. H Yoshimura, M Etoh, K Kondott and N Yokoya. Gray-Scale Character Recognition by Gabor Jets Projection, *in Proceeding 15th International Conference on Pattern Recognition*, 2:335-338, 2000.
23. R Battiti. First and Second Order Methods for Learning: Between Steepest Descent and Newton's Method, *in Neural Computation*, 4:141-161, 1992.
24. T Y Kwok and D Y Yeung. Constructive Algorithms for Structure Learning in Feedforward Neural Networks for Regression Problems, *in IEEE Transactions on Neural Networks*, 8:630-645, 1997.
25. Fahlman S E, Lebiere C. The Cascade-Correlation Learning Architecture, *in Advances in Neural Information Processing Systems*, 2:525-532, Morgan Kaufmann, 1989.
26. B Hassibi, D G Stork. Second Order Derivatives for Network Pruning: Optimal Brain Surgeon, *in Proceedings of Neural Information Processing System*, 5:164-171, 1992.
27. Sietsma J, R J F Dow. Creating Artificial Neural Networks that Generalize, *in Neural Networks*, 4(1):67-79, 1991.
28. M C Mozer, P Smolensky. Skeletonization: A Technique for Trimming the Fat from a Network via Relevance Assessment, *in Advances in Neural Information Processing Systems*, 1:107-115, Morgan Kaufmann, 1988.
29. E B Baum and D Haussler. What Size Net Gives Valid Generalization?, *in Neural Computation*, 1:151-160, 1989.
30. J E Moody. The Effective Number of Parameters: An Analysis of Generalization and Regularization in nonlinear Learning Systems, *In Advances in NIPS-4*, J.E. Moody editor, pages 847-854, Morgan Kaufmann, 1992.
31. P L Rosin, F Fierens. Improving Neural Network Generalization, *in Geoscience and Remote Sensing Symposium*, 2:1255-1257, 1995.

Dr. Siddhaling Urolagin has completed B.E in 2001 from Gogte Institute of Technology, Belgaum, M.Tech in 2004 and Ph.D in 2010 from Manipal Institute of Technology, Manipal, all in Computer Science and Engineering. He is currently working as Associate Professor in Department of Computer Science and Engineering at MIT, Manipal. His research interests include Gabor Filter based Feature Extraction, Artificial Neural Networks and Document Image Analysis. He is also carried out an AICTE funded project. He published several international/national journal and conference papers.

Dr. Prema K V, B.E, M.E, Ph.D. She is currently the Professor at Department of CSE, Faculty of Engineering and Technology, MITS University, Lakshmangarh. Her research interests inlude Artificial Intelligence, Pattern Recognition, Neural Networks, Computer Networks and Network Security. She has published around 25 papers in reputed journals/conferences. She is currently guiding 3 research scholars. She is also the principle investigator for one of the AICTE projects entitled "Design of an Intelligent System for Pattern Recognition using Soft Computing Models".

Dr. N V Subba Reddy. received Master of Engineering Degree, in Computer Science and Engineering, and PhD in Computer Science and Engineering from Indian Institute of Science, Bangalore and University of Mysore respectively. He is currently working as Vice Chancellor of Mody Institute of Technology and Science, Rajasthan. He was working as professor in Computer Science and Engineering Department and Associate Director Research and Devlopment, at Manipal Institute of Technology, Manipal University. He was also working as Director Sikkim Manipal Institute of Technology at Sikkim Manipal University, Sikkim. He has guided several Ph.D's in the area of Pattern Recognition, Neural Networks, Fuzzy Logic and Genetic Algorithms. Currently he is guiding Ph.D's in Document Image Processing, Character Recognition, and Machine Intelligence and Computer Cognition.