A NOVEL BACK PROPAGATION ALGORITHM WITH OPTIMAL NUMBER OF HIDDEN UNITS

Hiroaki Kurokawa, Chun-Ying Ho, Shinsaku Mori

Dept. of Electrical Engineering, Keio University, 3-14-1 Hiyoshi,
Kouhoku-ku, Yokohama, 223, JAPAN

Phone: +81 45 563 1141 Ext. 3319 FAX: +81 45 563 2773

email: kuro@mori.elec.keio.ac.jp

Abstract

A large fraction of recent work in artificial neural networks uses feed-forward networks trained with the back-propagation algorithm described by Rumelhart et.al. However, to decide the size of the network, how many layers and/or neurons, needed to solve a particular problem is a difficult question. This study discusses an approach to ease this problem based on hidden neuron's reduction method and shows an application for 3-layer networks. The network learned by proposed algorithm performs almost the same as the network with original back-propagation. By using this proposed method, the network structure (synaptic weight and number of neurons) can be decided simultaneously during a single learning.

summary

The multilayer feed-forward networks of interconnected of nonlinear neurons are widely used in adaptive problem [1]. One of the valuable characteristics of multilayer network is that the network can be trained. Back-propagation is a popular training algorithm and it applies to fields of technology like pattern recognition, classification etc. However, network performances are affected by the size of the network (layers and/or neurons). If training is started with too small network, no learning can occur and it needs to increase the size of the network and the network has to be retrained again. If the network is larger than that required, then learning is slowed, particularly in a conventional Von Neumann computer. The problem of the size of the network is a difficult problem to solve. Recent studies proposed some algorithms to avert these problems. The method based on information criteria was proposed in [2]. Although this method costs short time to determine the network size, the network that constructed by this method has no guarantee to converge to optimal solution. Also, the neuron reduction method was proposed in [3]. However, since the reduction of one neuron needs some epochs that are decided by human, a drastic time reduction is achieved for the determination of optimal number of hidden neurons. To solve this problem, we have proposed a new algorithm to optimize the size of the multilayer network under the learning with back-propagation in a single learning. This algorithm is based on the new character values C_j of hidden neurons which proposed here. These C_j s indicate neurons' stability and influence to the next layer. The value of C_j is given by,

$$C_j = \frac{2}{1-x_j^2} ,$$

where x_j is the output of the neuron. This equation is given by the differential form of the input-output function of the neuron (i.e. a sigmoid function which is bounded by (-1, 1)). This value C_j can be considered as how stable the output of the neuron is and how influential the output of the neuron is to the next layer. So it can be considered that neurons with large C_j are more important to the network than those with small C_j , then neurons which have a small C_j can be deleted. Since cutting down some neurons also means that the network loses its flexibility, these procedures are applied in the convergence stage of the learning. By cutting down neurons in the convergence stage, the influence of this matter can be almost negligible. We show simulation results of two standard test problems. First our proposed algorithm has been tested on artificial problem such as the XOR problem that is realized small size network. Then it has been tested on simple character classification as comparatively complex and large size network. These simulation results show that we can obtain an optimal size network even if this network size is not minimum size. And this optimized network performs as well as that initially structured to equal size network. This algorithm needs human's task. It is that we have to define the cutting line of neurons by the value of C_j . However this task is very easy to do, and is allowed some misjudgment. Therefore, since the number of final hidden neurons was concentrated around one value on both simulations XOR problem and character classification, it can be considered that the algorithm has flexibility for a complex problem.

References

- [1] D. E. Rumelhart, et al. : " Parallel Distributed Processing ", MIT Press (1986)
- [2] Takio Kurita. : " A Method to Determine the Number of Hidden Units of Three Layered Neural Networks by Information Criteria ", IEICE Trans. inf. & syst., Vol.J73-D-II, No.11, pp.1872-1878 (November 1990)
- [3] Masafumi Hagiwara. : " Back-Propagation with Artificial Selection — Reduction of the Number of Learning Times and That of Hidden Units — ", IEICE Trans. inf. & syst., Vol.J74-D-II, No.6, pp.812-818 (June 1991)