Fast Training MLP Networks with Lo-Shu Data Sampling

Hung Han Chen
Graphion
8787 Southside Blvd, Suite 503
Jacksonville, FL 32256
United States of America
hunghchen@mac.com

Abstract: Slow training speed is always one of the major concerns for MLP (Multi-Layer Perceptrons) networks. The remedies to this issue have been around by solving linear equations through the weights of hidden and output layers. By doing so, however, this solution will limit the usage of large networks and possibly large number of input features. Data sampling, on the other hand, has been used in statistics to speed up the modeling process. By introducing an ancient numeric concept, Lo-Shu, this paper presents a novel sampling technique to help training MLP networks potentially 3 times faster to generate acceptable models and still preserve the possibility of utilizing very large networks.

Key-Words: MLP, Artificial neural networks, Pruning, Fast training, Sampling, Lo-Shu

1 Introduction

MLP neural networks with Backpropagation learning algorithm have many issues since they have been suggested the potential of being a universal computing device [1]. Besides the issues of scaling up, local minima, and generalization [2][3], one major critic on MLP has been the slow speed of training.

Fast training neural networks have been studied by solving linear equations through the weights hidden and output layers [4]. However, the time needed for solving linear equations is increased exponentially as the increase of matrix dimensions. By doing so, this solution will limit the usage of large networks and possibly large number of input features. Therefore, in order to preserve the possibility of utilizing very large networks, other fast training methods need to be developed as well.

Sampling is a method used in statistics to choose samples from a data source. In the process of selecting samples from whole dataset, sampling reduces the resource, i.e., time, used on model building through reducing the number of training samples. The selected training samples, of course, need to, in some way, capture the real characterstics of the whole dataset.

As the issues for MLP networks are addressed one by one; scaling up in [5], local minima in [6], and generalization in [7], large MLP networks appear to generalize better than small networks since the operational condition of hidden neurons and the number of effective free parameters changes through the course of training. Therefore, the need of a fast training for large MLP networks remains. By introducing an ancient Chinese numeric concept, Lo-Shu, this paper presents a novel sampling technique to help training MLP networks potentially 3 times faster without sacrifying their computational simplicity.

The normal training curve of MLP networks is illustrated and discussed in Section 2. Automatic pruning [7] to help MLP modeling on generalization and speed is also mentioned in Section 2. The usage of Lo-Shu data sampling method is proposed in Section 3. Section 4 demonstrates the potential of Lo-Shu data sampling with simulations on a 4-class segmented image dataset. Conclusions and further studies are discussed in Section 5.

2 Training Curve of MLP networks

To make the MLP training useful, the learning algorithm needs to have the capability of solving complex problems, avoiding local minima, and generalizing well. Then the Mean Squared Error (MSE) will decrease rapidly at the beginning of training and converge to a minimum value through iterations. Fig. 1 shows a normal training curve for a MLP network.

It is well known that a larger MLP network can be trained to have a better result, compared to a smaller MLP network. But when the number of weights was considered as the number of free parameters on modeling process, most people are avoiding large networks in order to prevent overfitting. Some studies [7][8][9], however, have demonstrated that larger networks appear to generalize as well as smaller networks, sometimes