IDEA GROUP PUBLISHING
701 E Chocolate Avenue, Hershey PA17033, USA
Tel: 717/533-8845; Fax: 717/533-8661; URL: www.idea-group.com
ITB7260

Chapter XI

How to Train Multilayer Perceptrons Efficiently With Large Data Sets

Hyeyoung Park
Brain Science Institute, RIKEN, Japan

Feed forward neural networks or multilayer perceptrons have been successfully applied to a number of difficult and diverse applications by using the gradient descent learning method known as the error backpropagation algorithm. However, it is known that the backpropagation method is extremely slow in many cases mainly due to plateaus. In data mining, the data set is usually large and the slow learning speed of neural networks is a critical defect. In this chapter, we present an efficient on-line learning method called adaptive natural gradient learning. It can solve the plateau problems, and can be successfully applied to the learning associated with large data sets. We compare the presented method with various popular learning algorithms with the aim of improving the learning speed and discuss briefly the merits and defects of each method so that one can get some guidance as to the choice of the proper method for a given application. In addition, we also give a number of technical tips, which can be easily implemented with low computational cost and can sometimes make a remarkable improvement in the learning speed.

INTRODUCTION

Neural networks are one of the most important methodologies in the field of data mining and knowledge discovery, in which one needs to extract abstract knowledge such as rules or consistent properties from large data sets. With a large data set, one sometimes wants to find a functional mapping from input to output in order to give a prediction for a new input. On the other hand, one sometimes wants to cluster enormous amounts of data into several groups in order to get some information about the distributions of the data sets. Neural networks can manage such tasks well through the "learning process." Even though there are various neural network models, the multilayer perceptron (MLP) is one of the most popular models and so has been widely used for a variety of applications. In addition, it is also theoretically proven that MLP with one hidden layer can approximate any continuous function to any desired accuracy (Cybenko, 1989). In practical situations, however, those who are trying to use MLP as an application tool are often disappointed by the low learning performance such as poor generalization and slow convergence. In most cases, the low learning performance is caused by the use of a network with a bad structure or by the use of an inappropriate learning algorithm.

Following the development of the backpropagation learning algorithm (Rumelhart & McClelland, 1986), the MLP has been successfully applied to various fields of application such as pattern recognition, system control, time series prediction as well as data mining (Haykin, 1999; Michalski, Bratko, & Kubat, 1998). However, its slow learning speed has been a serious obstacle when used for real-world applications. Many studies have tried to solve this problem. Some of the studies are based on simple heuristics, while others are theoretical in nature. Even though there is little theoretical justification in the heuristic solutions, they are simple to apply and perform well in many cases. On the other hand, the theoretically inspired methods can give rigorous proof about their theoretical learning efficiency. However, they are sometimes too complex to apply to practical applications. Thus, it is important to select a method that is appropriate for an application in order to succeed.

In this chapter, we focus on the problem of slow convergence and propose a solution especially for learning using large data sets. We present an efficient learning method called adaptive natural gradient learning, which has been recently developed by Park, Amari, and Fukumizu (2000). We also give a number of technical tips for training MLP, which can be easily combined with general learning algorithms and the adaptive natural gradient learning method. In addition, we compare the adaptive natural gradient method with various popular learning algorithms for improving the learning speed and give short comments on the merits and defects of each method so that one can get some guidance towards choosing a proper method for a given problem.

In the Background section, we start our discussion by describing the structure of MLP and its learning process. In the section on How to Train Networks, we present simple tips for when designing MLP and explain various methods for