Chapter 5
On Simulation Performance of Feedforward and NARX Networks Under Different Numerical Training Algorithms

Salim Lahmiri
ESCA School of Management, Morocco

ABSTRACT

This chapter focuses on comparing the forecasting ability of the backpropagation neural network (BPNN) and the nonlinear autoregressive moving average with exogenous inputs (NARX) network trained with different algorithms; namely the quasi-Newton (Broyden-Fletcher-Goldfarb-Shanno, BFGS), conjugate gradient (Fletcher-Reeves update, Polak-Ribiére update, Powell-Beale restart), and Levenberg-Marquardt algorithm. Three synthetic signals are generated to conduct experiments. The simulation results showed that in general the NARX which is a dynamic system outperforms the popular BPNN. In addition, conjugate gradient algorithms provide better prediction accuracy than the Levenberg-Marquardt algorithm widely used in the literature in modeling exponential signal. However, the LM performed the best when used for forecasting the Moroccan and South African stock price indices under both the BPNN and NARX systems.

INTRODUCTION

Artificial neural networks are adaptive nonlinear systems capable to approximate any function. Theoretically, a neural network can approximate a continuous function to an arbitrary accuracy on any compact set (Funahashi, 1989; Hornik, 1991; Cybenko, 1989). The backpropagation (BP) algorithm that was introduced by Rumelhart (1986) is the well-known method for training a multilayer feed-forward artificial neural networks. It adopts the gradient descent algorithm. In the basic BP algorithm the weights are adjusted in the steepest descent direction (negative of the gradient). However, the backpropagation neural network (BPNN) has a slow learning convergent velocity and may be trapped in local minima. In addition, the performance of the BPNN depends on the learning rate parameter and the complexity

DOI: 10.4018/978-1-4666-8823-0.ch005

of the problem to be modelled. Indeed, the selection of the learning parameter affects the convergence of the BPNN and is usually determined by experience. Many faster algorithms were proposed to speed up the convergence of the BPNN. They fall into two main categories. The first category uses heuristic techniques developed from an analysis of the performance of the standard steepest descent algorithm. The second category uses standard numerical optimization techniques. The first category includes the gradient descent with adaptive learning rate, gradient descent with momentum, gradient descent with momentum and adaptive learning rate, and the resilient algorithm. In the standard steepest descent, the learning rate is fixed and its optimal value is always hard to find. The heuristic techniques allow the optimal learning rate to adaptively change during the training process as the algorithm moves across the performance surface. Therefore, the performance could be improved. The second category includes conjugate gradient, quasi-Newton, and Levenberg-Marquardt (LM) algorithm. In the conjugate gradient algorithms, a search is performed along conjugate directions; therefore the convergence is faster than steepest descent directions. Quasi-Netwon method often converges faster than conjugate gradient methods since it does not require calculation of second derivatives. For instance, it updates an approximate Hessian matrix at each iteration. Finally, The L-M method combines the best features of the Gauss-Newton technique and the steepest-descent method. It also converges faster than conjugate gradient methods since the Hessian Matrix is not computed but only approximated. For instance, it uses the Jacobian that requires less computation than the Hessian matrix.

BACKGROUND

In science and engineering problems, there are many papers in the literature that examined the effectiveness of each category of algorithms on the performance of the BPNN. For instance, Mokbnache and Boubakeur (2002) compared the performance of Levenberg-Marquardt, BP with momentum and BP with momentum and adaptive learning rate to classify the transformer oil dielectric and cooling state. They found that the BP with momentum and adaptive learning rate improves the accuracy of the BP with momentum and also gives a fast convergence to the network. Kisi and Uncuoglu (2005) compared Levenberg-Marquardt, conjugate gradient and resilient algorithm for stream-flow forecasting and determination of lateral stress in cohesionless soils. They found that Levenberg-Marquardt algorithm was faster and achieved better performance than the other algorithms in training. Esugasini et al. (2005) considered the problem of breast cancer diagnosis and compared the classification accuracy of the standard steepest descent against the classification accuracy of the gradient descent with momentum and adaptive learning, resilient BP, Quasi-Newton and Levenberg-Marquardt algorithm. The simulations show that the neural network using the Levenberg-Marquardt algorithm achieved the best classification performance. In their research, Iftikhar et al. (2008) employed three neural networks with different algorithms to the problem of intrusion detection in computer and network systems. The learning algorithms considered by the authors were the standard, the batch, and the resilient BP algorithm. They conclude that the resilient algorithm had a better performance to the application. Nouir et al. (2008) compared the performance of the standard BP with and Levenberg-Marquardt algorithm to the prediction of a radio network planning tool. They found that the standard BP algorithm achieved the minimum error and then outperforms the Levenberg-Marquardt algorithm. Recently, the performance of the BP trained with different algorithms to predict stock market trends has been examined in some recent works (Lahmiri, 2011; Lahmiri et al., 2014). It was found that the quasi-Newton numerical algorithm and Levenberg-Marquardt are the best in terms of accuracy.