Forecasting foreign exchange rates with an improved back-propagation learning algorithm with adaptive smoothing momentum terms

Lean YU (✉)1, Shouyang WANG1, Kin Keung LAI2
1 Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
2 Department of Management Sciences, City University of Hong Kong, Hong Kong, China

© Higher Education Press and Springer-Verlag 2009

Abstract The slow convergence of back-propagation neural network (BPNN) has become a challenge in data-mining and knowledge discovery applications due to the drawbacks of the gradient descent (GD) optimization method, which is widely adopted in BPNN learning. To solve this problem, some standard optimization techniques such as conjugate gradient and Newton method have been proposed to improve the convergence rate of BP learning algorithm. This paper presents a heuristic method that adds an adaptive smoothing momentum term to original BP learning algorithm to speedup the convergence. In this improved BP learning algorithm, adaptive smoothing technique is used to adjust the momentums of weight updating formula automatically in terms of “3 σ limits theory.” Using the adaptive smoothing momentum terms, the improved BP learning algorithm can make the network training and convergence process faster, and the network’s generalization performance stronger than the standard BP learning algorithm can do. In order to verify the effectiveness of the proposed BP learning algorithm, three typical foreign exchange rates, British pound (GBP), Euro (EUR), and Japanese yen (JPY), are chosen as the forecasting targets for illustration purpose. Experimental results from homogeneous algorithm comparisons reveal that the proposed BP learning algorithm outperforms the other comparable BP algorithms in performance and convergence rate. Furthermore, empirical results from heterogeneous model comparisons also show the effectiveness of the proposed BP learning algorithm.

Keywords back-propagation neural network, adaptive smoothing momentum, heuristic method, foreign exchange rates forecasting

1 Introduction

Since the revolutionary work of Rumelhart et al. [1] was published in 1986, artificial neural networks (ANNs) have been a common research stream in the past two decades. Over this time, the research stream has gain momentum with the advancement of computational technologies, which have made elaborate computation methods available and practically feasible [2]. At the same time, many different neural network models, such as back-propagation neural network (BPNN), radial basis function (RBF) neural network and self-organizing map (SOM) neural network, have already been developed. Among these models, the BPNN has been one of the most popular neural networks and has been widely applied to many fields, such as pattern recognition, function approximation, and system identification, due to its strong capability to approximate any arbitrary function arbitrarily well and to provide flexible nonlinear mapping between inputs and outputs [3, 4].

Basically, the generic learning rule of BPNN is based on the gradient descent (GD) optimization method and the chain rule [5]. However, some typical drawbacks of the BPNN learning rule based on GD are its slowness, its frequent confinement to local minima [6], and overfitting [7, 8]. For these reasons, some global optimization algorithms such as genetic algorithm (GA) and simulated annealing (SA) are proposed for escaping from local minima. Similarly, bias-variance trade-off [8, 9] and cross validation (e.g., k-fold
cross validation) [10] techniques are used to avoid overfitting problems. For the slow convergence rate of BP algorithm, there are several other ways to accelerate the BPNN learning.

The first way is the standard numerical optimization and nonlinear optimization techniques, such as quasi-Newton method, conjugate gradient, and Levenberg-Marquardt (LM) method. The crucial drawbacks of these methods, however, are that in many applications too much computation per pattern is required so that the storage and memory requirements go up as the square of the size of the network [11].

Another class of fast learning algorithms is the extended Kalman filter (EKF)-based techniques [12–14]. They have an improved convergence performance, but their numerical stability is not guaranteed and may degrade learning convergence and increase the learning time [15, 16].

Recently, a class of layer-by-layer optimizing algorithms [17, 18] was proposed to accelerate BP neural network learning. The basic idea of these methods is that each layer of the BPNN is decomposed into both a linear part and a nonlinear part, and the linear part of each layer is solved by way of the least squares method. These algorithms show fast convergence with reduced computational complexity relative to conjugate gradient algorithm. However, if the targets for the hidden layer cannot be linearly separated, then the mean squared error (MSE) cannot be sufficiently reduced at both the hidden layer and the output layer.

In such situations, a common method to overcome the slowness is to add a momentum term. As a heuristic method, the BP learning algorithm with a momentum term has been extensively reported in the past studies [19, 20]. Although introduction of the momentum may help dampen the oscillation of the weight adjustment, an improper choice of the momentum can actually slow the convergence rate. Furthermore, no clue has been given on how the momentum should be set. In most cases, it is determined empirically. In the existing literature, the momentum term is proportion to the previous weight increment. This way may increase network convergence speed if error change direction is consistent, but it may accelerate the network to be trapped into local minima because the weight adjustment (i.e., momentum) does not take the error change direction into consideration at all. To overcome this limitation, an adaptive smoothing momentum term is introduced into the weight update formulae of a BP learning algorithm based on adaptive smoothing technique and “3σ limits theory” [21, 22].

In this study, we propose an improved BP learning algorithm with adaptive smoothing momentum for foreign exchange rates forecasting. In this new algorithm, adaptive smoothing technique is used to adjust the momentum of weight updating formula automatically by tracking error signals in terms of “3σ limits theory”. For illustration and verification purposes, the proposed BP learning algorithm is applied to foreign exchange rates prediction.

The rest of this study is organized as follows: In Section 2, an improved BPNN learning algorithm with adaptive smoothing momentum terms is proposed in detail. In order to verify the effectiveness of the proposed algorithms, three typical foreign exchange rates, British pound (GBP), Euro (EUR), and Japanese yen (JPY), are chosen as the forecasting targets, and the corresponding computational results are reported in Section 3. Finally, some concluding remarks are drawn in Section 4.

## 2 Methodology formulation

In this section, an adaptive smoothing momentum is first introduced into weight updating formulas in terms of “3σ limits theory.” Then, an improved BP learning algorithm with adaptive smoothing momentum terms can be formulated.

### 2.1 Determination of adaptive smoothing momentum term

Consider a three-layer BP neural network, which has $p$ nodes in the input layer, $q$ nodes in the hidden layer, and $k$ nodes in the output layer. The main reason of using three-layer BP neural network reflects the following two fold. On one hand, three-layer BP neural networks are a class of typical BP neural network models, and many other BP networks can be obtained from this model. On the other hand, if some results can be obtained from three-layer BP neural network model, these results can be easily extended into other BP neural network models with different layers. Mathematically, the basic structure of the BPNN model [23] is described by

$$ y(t) = \left[ y_1(t + 1), y_2(t + 1), \ldots, y_k(t + 1) \right] $$

$$ f_2 \left[ \sum_{i=1}^{q} f_1 \left( \sum_{j=1}^{p} w_{ij}(t)x_j(t) + w_{i0}(t)\right)v_{i1}(t) + v_{i0}(t) \right] $$

$$ f_2 \left[ \sum_{i=1}^{q} f_1 \left( \sum_{j=1}^{p} w_{ij}(t)x_j(t) + w_{i0}(t)\right)v_{i2}(t) + v_{i0}(t) \right] $$

$$ \vdots $$

$$ f_2 \left[ \sum_{i=1}^{q} f_1 \left( \sum_{j=1}^{p} w_{ij}(t)x_j(t) + w_{i0}(t)\right)v_{ik}(t) + v_{i0}(t) \right] $$