Parameter by Parameter Algorithm for Multilayer Perceptrons

YANLAI LI, DAVID ZHANG and KUANQUAN WANG

Biocomputing Research Centre, School of Computer Science and Technology, Harbin Institute of Technology (HIT), Harbin, China.
Biometric Research Centre (UGC/CRC)/Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong. e-mail: csdzhang@comp.polyu.edu.hk

Abstract. This paper presents a parameter by parameter (PBP) algorithm for speeding up the training of multilayer perceptrons (MLP). This new algorithm uses an approach similar to that of the layer by layer (LBL) algorithm, taking into account the input errors of the output layer and hidden layer. The proposed PBP algorithm, however, is not burdened by the need to calculate the gradient of the error function. In each iteration step, the weights or thresholds can be optimized directly one by one with other variables fixed. Four classes of solution equations for parameters of networks are deducted. The effectiveness of the PBP algorithm is demonstrated using two benchmarks. In comparisons with the BP algorithm with momentum (BPM) and the conventional LBL algorithms, PBP obtains faster convergences and better simulation performances.

Key words. BP algorithm with momentum, layer by layer algorithm, multilayer perceptrons, parameter by parameter algorithm, training algorithm

Abbreviations. BPM – BP algorithm with momentum; LBL – Layer by Layer; MLP – Multilayer Perceptrons; MNN – Modular Neural Network; MSE – Mean Square Error; PBP – Parameter by Parameter

1. Introduction

Conventional techniques for training multilayer perceptrons (MLP) are mostly based on the method of gradient descent, that is, the weights update is performed in the opposite direction to the gradient of an error cost function [1]. Several methods were proposed for adapting the learning rates [2, 3]. Recently, Abid [4] modified the error function with a linear error term in the standard BP algorithm. Zweiri et al. [5] proposed a three-term algorithm to speed up the training procedure. These algorithms have a faster convergence in contrast with BP. Nevertheless, they are still based on the steepest descent method. One of the main drawbacks of this method is its slow rate of convergence, which limits its practical applications.

*Corresponding author.
The best-known algorithms for increasing the rate of convergence are the conjugate training algorithm and the Levenberg Maquardt (LM) training algorithm [6, 7]. Both have the character of quadratic convergences, yet this type of algorithms is not useful for large-scale problems because the memory requirement and the computational complexity increase quadratically according to the number of network parameters.

Least squares based algorithms (LSB) have been proposed to accelerate the training process by separating each layer into linear and nonlinear parts. Each linear part is optimized separately by using a least square criterion [8, 9]. This method has the advantage of decreasing network error to a small value after the first few iterations [9, 10]. Unfortunately, it often encounters difficulties in the continuing calculation.

The layer by layer (LBL) optimizing algorithm is also based on the least squares method and constrained optimization [11, 12]. This algorithm shows fast convergence with much less computational complexity than either the conjugate gradient or Newton methods. Unfortunately, however, it has a stalling problem. When the hidden targets become linearly inseparable, it is impossible to reduce the error sufficiently [13]. Several methods [11, 14], were proposed to solve the stalling problem, all based on heuristic rules. Oh and Lee [13] have proposed a new error function at hidden layers to overcome the stalling problem without heuristics. With this new hidden error function, the LBL algorithm approximately converges with the error backpropagation algorithm with optimum learning rates.

In this paper, a new parameter by parameter (PBP) algorithm is proposed on the basis of the LBL idea. In each iteration epoch, the PBP algorithm optimizes the weights and thresholds directly, one by one with other variables fixed. Calculations concentrate only on the changed part, greatly decreasing the computation complexity. Furthermore, we choose the Moore-Penrose inverse method [15] to solve the stalling problem inherent in the LBL algorithm.

This paper is organized as follows. Section 2 introduces the PBP algorithm in detail. Section 3 discusses the PBP algorithm stalling problem and possible solutions. Some theoretic analysis about the proposed algorithm is given in Section 4. Section 5 shows the experimental results for PBP algorithm. Finally, Section 6 provides a brief conclusion to this paper.

2. The Parameter by Parameter Algorithm

Since the property of each hidden layer in a general MLP has similarity, let us consider, for the sake of simplicity, a fully-connected network with only one hidden layer. Suppose that the number of input, hidden, and output units are $M$, $H$, $N$, respectively. In addition, the number of training patterns is $P$. In the batch mode case, the error function of an output layer can be written as: