East Asian Journal on Applied Mathematics
doi: 10.4208/eajam.2023-025.080823

Vol. **14**, No. 4, pp. 769-787
November 2024

Boundedness and Convergence Analysis of a Pi-Sigma Neural Network Based on Online Gradient Method and Sparse Optimization

Qinwei Fan[1,2,*], Le Liu[1], Shuai Zhao[1], Zhiwen Zhang[1], Xiaofei Yang[1], Zhiwei Xing[1] and Xingshi He[1]

[1]*School of Science, Xi'an Polytechnic University, Xi'an 710048, China.*
[2]*School of Mathematics and Information Science, Guangzhou University, Guangzhou, 510006, China.*

Received 18 January 2023; Accepted (in revised version) 8 August 2023.

Abstract. High order neural networks have strong nonlinear mapping ability, but the network structure is more complex, which restricts the efficiency of the network, and the relevant theoretical analysis is still not perfect up to now. To solve these problems, an online gradient learning algorithm model of Pi-Sigma neural network with a smooth set lasso regular term is proposed. Since the original lasso regular term contains absolute values and is not differentiable at the origin, it causes experiment oscillations and poses a great challenge to the convergence analysis of the algorithm. We use grinding technology to overcome this deficiency. The main contribution of this paper lies in the adoption of online learning algorithm, which effectively improves the efficiency of the algorithm. At the same time, strict theoretical proofs are presented, including strong convergence and weak convergence. Finally, the effectiveness of the algorithm and the correctness of the theoretical results are verified by numerical experiments.

AMS subject classifications: 65M10, 78A48

Key words: Online gradient method, Pi-Sigma neural network, regularizer, convergence.

1. Introduction

As an important type of high order neural networks, Pi-Sigma neural network has high learning efficiency, strong robustness, and powerful nonlinear processing ability. Therefore, such networks have attracted wide attention and are widely used in various fields [2, 7–9, 12, 27, 37, 38].

Backpropagation (BP) algorithm is the most popular method in supervised training of feedforward neural networks. It takes the form of minimizing the mean square error between the expected response and the actual response [11, 25, 29, 31]. In theory, a network

*Corresponding author. *Email addresses:* qinweifan@xpu.edu.cn (Q. Fan), lliu737784717@163.com (L. Liu), z402633008@126.com (S. Zhao), zhiwenzhang0316@126.com (Z. Zhang), yangxiaofei2002@163.com (X. Yang), zwxing@xpu.edu.cn (Z. Xing), xsh1002@126.com (X. He)

with enough neurons can approximate any function with any accuracy, but determining a reasonable structure of the neural network is still a challenging problem.

The gradient method is a commonly used neural network training method to minimize the error function. This can be achieved by a batch gradient method or an online gradient method [19,33,42]. Specifically, batch learning algorithms update weights only once after all samples are presented to the network [6, 20, 32]. Online learning algorithms, on the other hand, update the weights every time a sample is presented to the network [13].

According to the difference of input form of the sample points, it can be divided into online gradient algorithm based on sequential input samples, online gradient algorithm based on specific random input samples and online gradient algorithm based on completely random input samples, including random more strong more conducive to jump out of local minimum, however, due to the introduction of randomness, it becomes challenging to analyze the theoretical performance of the algorithm [16, 34]. In the recent years, there are some theories and applications of gradient-based neural networks have been reported. In [41], a novel finite-time convergent gradient-based neural network model is proposed for solving the dynamic Moore-Penrose inverses problem, and its finite-time convergence is preserved even in the presence of additive bounded dynamic noises. Also, a unified gradient-based neural networks model is proposed for both static matrix inversion and time-varying matrix inversion with finite-time convergence regardless of the existence of bounded additive noises [40].

Generally speaking, the generalization effect of neural networks with smaller weights is better [1]. Two main aspects of neural network learning consist in preventing the over fitting caused by excessive weights and eliminating unnecessary weight connections to achieve a sparse network structure. To solve these problems and optimize the network, one often uses a simplified model, early stop, data enhancement and regularization.

As we all know, adding regularization term to traditional error function can effectively sparsely optimize network structure and obtain better generalization performance [3,4,15]. The error function with the regularization term is as follows:

$$\text{Error} = \frac{1}{2}\sum_i \|\text{Output}^i - \text{Target}^i\| + \lambda \ell(W),$$

where parameter $\lambda > 0$ is the regularization coefficient.

Here we introduce several common regularization terms, L_0 regularization produces the most sparse solution, but it is not easy to calculate [21, 35]. On the other hand, L_1 regularization can produce sparse weight matrices — i.e. sparse models that can be used for feature selection [24]. At the same time, the L_1 regularizer is the optimal convex approximation of L_0 regularizer. Unfortunately, they cannot sparsely select weights at the group level, and both of them are NP-hard problems and are not easy to solve. L_2 regularization can effectively inhibit the excessive growth of weights and prevent the model from overfitting [18, 26, 39, 42]. But L_2 regularization does not any have sparsity.

As a compromise between L_0 and L_1 regular term, a regularization method of $L_{1/2}$ is proposed. However, the regularization term is not differentiable at the origin of coordinates, which makes theoretical analysis difficult. To overcome this defect, a smoothing technique is proposed [5, 14, 19, 23].