Neurocomputing 72 (2009) 2979–2999

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Iterative algorithm of wavelet network learning from nonuniform data

Zhiguo Zhang *

University of Electronic Science and Technology of China, College of Automation, Jianshe North Road, Chengdu, Sichuan 610054, China

A R T I C L E I N F O

Article history:
Received 13 October 2007
Received in revised form
2 August 2008
Accepted 17 April 2009
Communicated by T. Heskes
Available online 14 May 2009

Keywords:
Convergence of algorithm
Wavelet neural networks
Nonuniform sampling

A B S T R A C T

The learning algorithm based on multiresolution analysis (LAMA) is a powerful tool for wavelet networks. It has many advantages over other algorithms, but it seldom does well in the learning of nonuniform data. A new algorithm is proposed to solve this problem, which develops from the learning algorithm based on sampling theory (LAST). From the good concentration of wavelet energy, we discuss the approximation capacity of wavelet network in the local domain when the training data are not dense enough. From this discussion, the new algorithm is realized by the iterative application of LAST. The corresponding theorems based on the sampling theory are also proposed to prove the rationality of new algorithm. In the simulation, we compare the performance of new algorithm with that of LAMA and LAST. The results show that our new algorithm has as many advantages as LAMA and LAST, does better in the learning of nonuniform data and has high approximation accuracy.

1. Introduction

Neural networks are found to be an effective way to identify the nonlinear dynamic systems with incomplete knowledge or even no knowledge. But, the algorithms of neural networks always suffer from slow convergence, local minimum and overfitting. Recently, the excellent properties of wavelets have inspired people to solve these problems by wavelet networks. Usually, there are two kinds of wavelet networks, which are constructed from different ideas [1]. One is to take the wavelet network as a special RBF network [2,3], whose algorithm is to some extent similar to that of RBF networks. The other is constructed from the multiresolution analysis (MRA) (multiresolution approximation) [4].

Whichever wavelet network is taken, the error back propagation may be the most popular algorithm in the learning [2–4] and is often combined with orthogonal least square-backward elimination (OLS-BE) [2,3], which is used in the selection of network structures. Though the error back propagation is improved by the introduction of wavelets, it still cannot avoid local minimum and overfitting completely. In order to deal with these problems, many limitations are imposed on the parameters of wavelet networks when the error back propagation and OLS-BE are applied [2,3]. However, these limitations often force the error back propagation to adjust so many parameters during training [5,6] that algorithm becomes very complex and requires huge computations.

Of course, some algorithms are also proposed to mitigate the shortcoming of the error back propagation. By fully considering the smoothness and the initialization of weights [14,15] or incorporating the fuzzy theory and the Kalman filter etc. into the algorithm [5,6], just as Ho has done [7], the error back propagation is improved greatly. However, though these attached algorithms have obtained some successes in the error back propagation, they still cannot help wavelet network get rid of its troubles totally.

Due to the ineffectiveness of error back propagation, many works also try to train wavelet networks from the view point of evolutionary algorithms [8], where the evolutionary algorithms are used first to locate a good region in the parameter space and then the gradient descent algorithm, the local search procedure, is adopted to determine a near optimal solution in that region [9,10]. However, though the evolutionary algorithms are global searching methods, most evolutionary algorithms are rather inefficient and cannot avoid certain degeneracy and local minimum completely.

The algorithms above have accelerated convergence, avoided local minimum and overcome overfitting in some extent, but they still do not solve the problems of wavelet network completely. This mostly comes from the fact that these algorithms stem from that of typical neural networks, so they seldom utilize fully the excellent properties of wavelets in the frequency-domain.

Instead of typical algorithm of neural network, Zhang proposes an learning algorithm based on multiresolution analysis (LAMA) [4]. Since the input weights and the output weights have the different meanings for the wavelet networks, i.e. the input weights determine the approximation space (multiresolution space) and the output weights represent the coefficients of function basis in the approximation space, LAMA adjusts input weights and output weights in the different methods [4]. Due to this disposal, the algorithm of Zhang accelerates the convergence, avoided local minimum and overcome overfitting effectively, so it

* Tel.: +86 15928960590.
E-mail address: zhangzg2002@hit.edu.cn

0925-2312/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.04.014

is soon used, widely, in the wavelet network constructed from the multiresolution analysis [6,11,12]. However, there still exist some problems in his algorithm.

One of them is that the algorithm in [4] is designed under the assumption that the domain of interest can be divided into several parts where the training data more or less uniform. This disposal has simplified the learning of wavelet network, but limits the application of algorithm since the distribution of training data is not always uniform.

The learning of wavelet network from nonuniform data is very similar to the problem of nonuniform sampling, where a signal is reconstructed from a finite number of its unevenly spaced sampled data.

A famous theorem of nonuniform sampling is Kadec's $\frac{1}{4}$-theorem, which states that the signal can be reconstructed from nonuniform data just as from uniform data if the perturbation of sampling period is not more than $\frac{1}{4}$ sampling period [17,18]. This theorem has influenced the development of nonuniform sampling theorem [20,21] and forms the basis of typical nonuniform sampling theorem such as Lagrange-type interpolation [22].

Another nonuniform sampling theory is the jittered sampling. By jittered samples, we mean the nonuniform samples that are clustered around uniform. Many algorithms of jittered sampling are developed on the assumption that there exists a one-to-one mapping from a nonuniform data set to the uniform one [23,35]. By this mapping, the Fourier transform is analyzed and the aliasing from the nonuniformity of samples is avoided in low frequency band [25]. In addition, there are many other algorithms of jittered samples which are not related to such a mapping. Many of them are proposed by some typical technologies such as a Cartesian grid [24], a weighted algorithm [25] and a probability distribution [27].

Period nonuniform sampling (Papoulis' generalized sampling) is another important nonuniform sampling theorem, which has the close connection with the quadrature mirror filters [28,29]. The periodic nonuniform sampling assume that the nonuniform data are formed by the discretization of the outputs from several parallel filters at uniform sampling period after the input of signal into these filters. Papoulis' generalized sampling is used widely in many actual cases, so it is soon extended to a multiple-input multiple-output sampling scheme and many algorithms are proposed to analyze its spectral properties and mitigate its disadvantages [26,28,29].

Since both periodic nonuniform sampling and wavelet are closely connected with quadrature mirror filters, wavelets are also used in the interpolation of nonuniform samples [19,32]. Nonuniform interpolation by wavelet and scaling function not only effectively reconstructs the signals by the global information, but also fully uses the locally dense samples to reconstruct at higher resolutions [30].

The nonuniform sampling theorems have made great progress, but few of them can be applied to the learning of wavelet network directly. For examples, the preconditions of some sampling theorems, such as period nonuniform sampling, are unavailable for wavelet network and many of them often do not work well in the interesting domain where there exist some regions containing sparse data [23,35], which is just what the neural network has to deal with. The trouble of nonuniform sampling theorems mostly arises from the fact that, for some nonuniform distribution, they do not have such an effective method to describe the information of target function in discrete data as Shannon theorem, which has described this by the Fourier transform of discrete data.

Many works also devote to the Fourier transform of the nonuniform Dirac comb or data in the different models [33,36] and propose the algorithm to remove the spectral bias of nonuniform Fourier transform [23,34], but the influence of nonuniformity on the Fourier transform is still difficult to estimate for some nonuniform distributions.

Though many typical sampling is difficultly applied to the learning of neural network directly, some of them are very suggestive for the learning of neural network. They indicate that some nonuniform distributions of data can be considered as the uniform [20,21], many nonuniform data sets can be measured by several uniform data sets [26,28,29] and the approximation capacity of wavelets in different resolutions is very suitable for nonuniform interpolation [30–32].

The combination of wavelet and neural network has supplied an access to absorb these viewpoints in the learning of wavelet network. In our previous work [13], we have connected the sampling theorem to the learning of wavelet network from uniform data. Here, from the viewpoints of nonuniform sampling, a novel algorithm, which is closely connected with the localization of wavelet energy in the time–frequency plane, is proposed to train wavelet network on the nonuniform data by the iterative application of the algorithm in [13]. We show that this algorithm not only has the capacity to learn from the nonuniform data which are more or less uniform, but also are suitable for the learning of wavelet network in the domain where there are some regions containing sparse data.

This work is divided into four parts. In Section 2, the first part, we briefly review some important theoretical results relative to our algorithm. Then, in the second part consisting of the Section 3, the iterative learning algorithm based on sampling theory (ILAST), which develops from learning algorithm based on sampling theory (LAST), is proposed to learn from the nonuniform data. In this part, we discuss two iterative procedures, which correspond, respectively, to the good localization of activation function energy in the frequency-domain and the time–frequency plane. Section 4, the third part, discusses the rationality of ILAST. In the last part, the simulation shows that our new algorithm learns from the nonuniform data with the high accuracy and avoids the overfitting effectively.

2. Background

We will use standard notation throughout, which is listed in Appendix D. In our work, the wavelet network is constructed from a multiresolution analysis and the Fourier transform is with respect to any an input variable instead of only time.

For an effective explanation, though there does not exist a function of compact support whose Fourier transform has a compact support, in our algorithm, we take approximately a scaling function or wavelet as the function such that both its Fourier transform and itself have compact supports because of its good localization of energy in time–frequency plane.

2.1. Wavelet network in our algorithm

In this subsection, we describe the wavelet network which is discussed in our algorithm. For more details of its structure and presentation, please refer to [4].

The wavelet network with three layers is introduced in [4], which has the following presentation:

$$f_{ne}(x) = \sum_{k=-\infty}^{\infty} c_{J,k}\phi(2^J x - k) = \sum_{k=-\infty}^{\infty} c_{J,k}\phi_{J,k}(x). \tag{1}$$

Consider the compact interval of interest. Formula (1) is also written as

$$f_{ne}(x) = \sum_{k=I_0}^{I_1} c_{J,k}\phi_{J,k}(x) \tag{2}$$