Neurocomputing 72 (2009) 1808–1816

Contents lists available at ScienceDirect

Neurocomputing

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

Functional-link nets with genetic-algorithm-based learning for robust nonlinear interval regression analysis

Yi-Chung Hu *

Department of Business Administration, Chung Yuan Christian University, Chung-Li 32023, Taiwan, ROC

ARTICLE INFO

Article history:
Received 7 December 2007
Received in revised form
30 April 2008
Accepted 2 July 2008
Communicated by S. Mitra
Available online 24 July 2008

Keywords:
Functional-link net
Interval regression analysis
Genetic algorithm
Fuzzy regression
Robust learning algorithm

ABSTRACT

Interval regression analysis has been a useful tool for dealing with uncertain and imprecise data. Since the available data often contain outliers, robust methods for interval regression analysis are necessary. This paper proposes a genetic-algorithm-based method for determining two functional-link nets for the robust nonlinear interval regression model: one for identifying the upper bound of data interval, and the other for identifying the lower bound of data interval. To facilitate the inclusion of regular data in the robust nonlinear interval regression model, in the fitness function, not only the cost function with different weighting schemes but also the number of training data included in the interval model is taken into account. As for resisting outliers, the effects of training data beyond or beneath the estimated data interval on the determination of upper and lower bounds can be greatly reduced during the training phase when these data are located in the rejection region. Simulation results demonstrate that the proposed method performs well for contaminated data sets by resisting outliers and including all regular data in the data intervals.

1. Introduction

In many practical applications, since the available information is often derived from uncertain assessments, real intervals can be employed to represent uncertain and imprecise observations [8]. Interval regression analysis, which provides an interval estimation of individual dependent variables, is an important tool for dealing with uncertain data [14,8,7]. The interval parameters of a linear interval model can be determined by solving a basic linear programming problem of interval regression analysis [9,12]. Interval regression analysis was developed on the basis of an important tool, namely fuzzy regression analysis introduced by Tanaka et al. [23], whose objective is to build a model that contains all observed output data included in the system in terms of fuzzy numbers [23,25]. Fuzzy regression analysis has been successfully employed in different applications such as market forecasting [6], identification [15], house price estimation [24], quality evaluation [1], and so on.

In view of the capability of neural networks as an approximator of nonlinear mappings, many neural-network-based approaches have been proposed for fuzzy regression analysis. For instance, radial basis function networks were considered by Cheng and Lee [3], in which the predefinition of functional relationship between the input and the output was not required. Ishibuchi et al. [11] proposed an architecture of neural networks for interval target values. Later, asymmetric fuzzy coefficients and fuzzified neural networks were put forward by Ishibuchi and Nii [12]. As for nonlinear interval regression analysis, Ishibuchi and Tanaka [9] employed two backpropagation (BP) multi-layer perceptrons (MLPs) to represent the upper and lower bounds of data interval by the least square error from a given data set. Nevertheless, the inclusion of given data in a data interval cannot be assured by this method. That is, the data interval determined by two MLPs approximately includes all given training data.

When training data are not contaminated by outliers, the above-mentioned methods perform well by including almost all given training data in the data interval. Nevertheless, since training data are often corrupted by outliers, data interval obtained by these methods may be influenced by outliers. The robust nonlinear interval model for reducing the effects of outliers on the interval regression analysis has been an interesting area of research, whereas neural networks have been effective tools for identifying the upper and lower bounds of data interval. For instance, Huang et al. [7] employed two MLPs to determine nonlinear interval models using a new cost function, in which the cost function in [9] and the robust BP algorithm for function approximation [2] were taken into account. Jeng et al. [14] proposed the support vector interval regression networks consisting of two radial basis function networks to determine the upper and lower bounds. Moreover, Hwang et al. [8] proposed a

* Tel.: +886 3 2655130; fax: +886 3 2655199.
E-mail address: ychu@cycu.edu.tw

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.07.002

robust method by combining the possibility estimation formulation integrating the property of central tendency with the principle of support vector interval regression.

It is found that although the above-mentioned robust learning algorithms for determining nonlinear interval models are robust against outliers, it seems that not all regular data can be included in the estimated data interval by these methods. This paper aims to propose a genetic-algorithm-based (GA-based) method for determining the robust nonlinear interval regression model, which can facilitate the inclusion of all regular data in the estimated data interval. The proposed robust learning algorithms have two characteristics: one is that not only the cost function with different weighting schemes proposed in [9], but also the number of training data included in the interval model are incorporated into the fitness function, and the other is that the idea of the quality of training patterns introduced in [7,2] is taken into account. In practice, the upper limit on the percentage of outliers beyond and beneath the true interval model can be specified so as to identify the rejection region. The training data beyond or beneath the estimated data interval would generate very little influence on the determination of both upper and lower bounds when they are located in the rejection region. In particular, regarding the neural model, owing to the simple architecture and the function approximation capability of the functional-link net [18–20], it is interesting to examine the feasibility of employing two functional-link nets to identify the upper and lower bounds of data interval. To sum up, the main contribution of this paper is to propose a new GA-based method for constructing a robust nonlinear interval model.

The rest of this paper is organized as follows: the functional-link net with the functional-expansion model for approximation is introduced in Section 2. Section 3 introduces the MLP-based approach proposed by Ishibuchi and Tanaka [9] for nonlinear interval regression analysis. Section 3 describes the proposed GA-based learning algorithms in detail. In Section 4, in order to examine the effectiveness and applicability of the proposed method for determining a nonlinear interval regression model, three examples and two real data are taken into account. Discussion and conclusions are presented in Section 5.

2. Neural network models for nonlinear interval regression analysis

In this section, since the proposed GA-based method employs two functional-link nets to determine the estimated data interval, and the weighting schemes introduced in [9] are incorporated into the proposed method, the functional-link net and the MLP-based approach proposed by Ishibuchi and Tanaka are introduced in Sections 2.1 and 2.2, respectively. Previously, Chen and Jain [2] employed time-dependent cutoff points to determine the degree of influence of each training pattern on the training process for a robust backpropagation algorithm for function approximation. Since time-dependent cutoff points play an important role in reflecting the quality of the training data for robust learning algorithms [7], they are introduced in Sections 2.3.

2.1. Functional-link net

A functional-link net is a one-layer feed-forward network. The sigmoid function $f_h(u) = 1/(1+e^{-u})$ is commonly used as the transfer function for the output node. The input pattern of a functional-link net can either be a functional-expansion representation, tensor representation or a combination of these two representations. Pao [18] demonstrated the effectiveness of the functional-expansion representation using a set of orthogonal functions in function approximation. In other words, the functional-link net with a functional-expansion model can be used as a tool for the approximation of nonlinear functions. For instance, the functional link in the functional-expansion model can be designed to generate $\{x, \sin(\pi x), \cos(\pi x), \sin(2\pi x), \cos(2\pi x), \ldots\}$ as an enhanced representation for a single input x. In addition, Pao [19] pointed out that the tensor model could be ineffective for function approximation.

Let us denote the given non-fuzzy input–output pairs by $(\boldsymbol{x}_p, y_p)$, $p = 1, 2, \ldots, m$, where $\boldsymbol{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$ and y_p are input vector and the corresponding desired output value, respectively. For simplicity, the learning of a function of one variable is taken into account. Without losing generality, let $\boldsymbol{x}_p$ be represented by (x_p). Regarding the functional-link net with the functional-expansion model, an enhanced input vector, $(x_p, \sin(\pi x_p), \cos(\pi x_p), \sin(2\pi x_p), \cos(2\pi x_p))$ is obtained by the functional link, and is presented to the functional-link net subsequently. Let θ be the bias to the output node. Then, the actual output value o_p corresponding to $(x_p, \sin(\pi x_p), \cos(\pi x_p), \sin(2\pi x_p), \cos(2\pi x_p))$ is calculated as follows:

$$o_p = f_h(w_1 x_p + w_2 \sin(\pi x_p) + w_3 \cos(\pi x_p) + w_4 \sin(2\pi x_p) + w_5 \cos(2\pi x_p) + \theta) \tag{1}$$

To train the functional-link net, the following sum square cost function is used:

$$E = \frac{1}{2}\sum_{i=1}^{m}(y_p - o_p)^2 \tag{2}$$

where m is the number of training patterns. Both y_p and o_p range between 0 and 1.

2.2. MLP-based approach

In view of the high capability of neural networks for nonlinear regression, Ishibuchi and Tanaka [9] employed two MLPs, MLP* and MLP$_*$, to enhance the usefulness of the interval regression analysis. The idea is that a nonlinear interval model can be derived from two nonlinear functions. Each of the two networks has n inputs, a single output and only one hidden layer. Let $g^*(\boldsymbol{x})$ and $g_*(x)$ denote the output functions realized by MLP* and MLP$_*$, respectively. In practice, $g^*(\boldsymbol{x})$ and $g_*(x)$ represent the upper and lower bounds of a nonlinear interval model, respectively.

A nonlinear optimization problem is formulated to determine the nonlinear interval regression model as follows:

$$\text{Minimize } (g^*(\boldsymbol{x}_1) - g_*(\boldsymbol{x}_1)) + (g^*(\boldsymbol{x}_2) - g_*(\boldsymbol{x}_2)) + \cdots + (g^*(\boldsymbol{x}_m) - g_*(\boldsymbol{x}_m)) \tag{3}$$

$$\text{subject to } g_*(\boldsymbol{x}_p)) \leqslant y_p \leqslant g^*(\boldsymbol{x}_p), \quad p = 1, 2, \ldots, m \tag{4}$$

where $(g^*(\boldsymbol{x}_p) - g_*(\boldsymbol{x}_p))$ represents the width of the estimated data interval for $\boldsymbol{x}_p$. The objective of the above formulation is to determine the nonlinear interval model with the least sum of the widths of the predicted intervals for the respective inputs subject to that estimated data interval determined by the two MLPs including all the given input–output pairs.

Instead of deriving a learning algorithm directly from the above nonlinear optimization problem, Ishibuchi and Tanaka derived learning algorithms for $g^*(\boldsymbol{x})$ and $g_*(x)$ by the following cost function with the weighting scheme ω_p:

$$E = \sum_{p=1}^{m}\frac{1}{2}\omega_p(y_p - g^*(\boldsymbol{x}_p))^2 \tag{5}$$