Information Sciences 412–413 (2017) 210–222

Contents lists available at ScienceDirect

Information Sciences

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

Robust stochastic configuration networks with kernel density estimation for uncertain data regression

Dianhui Wang*, Ming Li

Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3086, Australia

ARTICLE INFO

Article history:
Received 16 February 2017
Revised 29 May 2017
Accepted 30 May 2017
Available online 31 May 2017

Keywords:
Stochastic configuration networks
Robust data regression
Randomized algorithms
Kernel density estimation
Alternating optimization techniques

ABSTRACT

Neural networks have been widely used as predictive models to fit data distribution, and they could be implemented through learning a collection of samples. In many applications, however, the given dataset may contain noisy samples or outliers which may result in a poor learner model in terms of generalization. This paper contributes to a development of robust stochastic configuration networks (RSCNs) for resolving uncertain data regression problems. RSCNs are built on original stochastic configuration networks with weighted least squares method for evaluating the output weights, and the input weights and biases are incrementally and randomly generated by satisfying with a set of inequality constrains. The kernel density estimation (KDE) method is employed to set the penalty weights for each training samples, so that some negative impacts, caused by noisy data or outliers, on the resulting learner model can be reduced. The alternating optimization technique is applied for updating a RSCN model with improved penalty weights computed from the kernel density estimation function. Performance evaluation is carried out by a function approximation, four benchmark datasets and a case study on engineering application. Comparisons to other robust randomised neural modelling techniques, including the probabilistic robust learning algorithm for neural networks with random weights and improved RVFL networks, indicate that the proposed RSCNs with KDE perform favourably and demonstrate good potential for real-world applications.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

For many real-world applications, sample data collected from various sensors may be contaminated by some noises or outliers [7], which makes troubles for building neural networks with sound generalization. Over the past years, robust data modelling techniques have received considerable attention in the field of applied statistics [7,8,11] and machine learning [2,3,5,12,19]. It is well known that the cost function plays an important role in robust data modelling. In [2], the M-estimator and Hampels hyperbolic tangent estimates were employed in the cost function, aiming to alleviate the negative impacts of outliers on the modelling performance. Under an assumption that the additive noise of the output follows Cauchy distribution, the mean log squared error was used as a cost function in [12]. In [5], a robust learning algorithm based on the M-estimator cost function with random sample consensus was proposed to deal with outliers, and this algorithm has been successfully applied in computer vision and image processing [15,20,22]. Besides these methods mentioned above, some results on robust data regression using support vector machine (SVM) have been reported in [3,19], where SVM-based approaches demonstrate some limits to handle uncertain data regression problems with higher level outliers.

* Corresponding author.
E-mail address: dh.wang@latrobe.edu.au (D. Wang).

http://dx.doi.org/10.1016/j.ins.2017.05.047
0020-0255/© 2017 Elsevier Inc. All rights reserved.

Back-propagation algorithms for training neural networks suffer from many shortcomings, such as learning parameter setting, slow convergence and local minima. Thus, it is useful to develop advanced learning techniques for resolving data regression problems, in particular, for stream data or online data modelling tasks. With such a background, randomized methods for training neural networks have been developed in the last decades [9,14,16]. Readers may refer to a recently published survey paper for more details about some milestones on this topic [18]. Studies on the robust data modelling techniques based on Random Vector Functional-link (RVFL) networks have been reported in [1,4]. Specifically, a hybrid regularization model with assumption on the sparsity of outliers was used in training process, and a probabilistic robust learning algorithm for neural networks with random weights (PRNNRW) was proposed in [1]. However, some learning parameters used in PRNNRW must be set properly and this is quite difficult to be done in practice. In [4], an improved version of RVFL networks built by using a KDE-based weighted cost function was suggested. Unfortunately, the significance of the scope setting of the random weights and biases for RVFL networks has not been addressed. In [13], we looked into some practical issues and common pitfalls of RVFL networks, and clearly revealed the impact of the scope setting on the modelling performance of RVFL networks. Our findings reported in [13] motivates us to further investigate the robust data regression problem using an advanced randomized learner model, termed as Stochastic Configuration Networks (SCNs), which are built incrementally by assigning the random weights and biases with a supervisory mechanism [21].

This paper aims to develop a robust version of SCNs for uncertain data regression. Based on the construction process of SCNs, we utilise a weighted least squares objective function for evaluating the output weights of SCNs, and the resulting approximation errors from the present SCN model are used to incrementally configure the hidden nodes with constrained random parameters. During the course of building RSCNs, the penalty weights representing the degree of contribution of individual data samples to the objective function are updated according to a newly constructed KDE function. In this work, an alternating optimization (AO) technique is employed to implement the RSCN model. Our proposed algorithm, termed as RSC-KDE, is evaluated by using a function approximation, four benchmark datasets with different levels of artificial outliers, and an engineering application [4]. Experimental results indicate that the proposed RSC-KDE outperforms other existing methods in terms of effectiveness and robustness.

The remainder of paper is organized as follows: Section 2 briefly reviews stochastic configuration networks and the kernel density estimation method. Section 3 details our proposed RSC-KDE algorithm. Section 4 reports experimental results with comparisons and discussion. Section 5 concludes this paper with some remarks.

2. Revisit of stochastic configuration networks

This section reviews our proposed SCN framework, in which the input weights and biases are randomly assigned in the light of a supervisory mechanism, and the output weights are evaluated by solving a linear least squares problem. More details about SCNs can be read in [21].

Let $\Gamma := \{g_1, g_2, g_3 \ldots\}$ be a set of real-valued functions, span(Γ) denote a function space spanned by Γ; $L_2(D)$ denote the space of all Lebesgue measurable functions $f = [f_1, f_2, \ldots, f_m] : \mathbb{R}^d \to \mathbb{R}^m$ defined on $D \subset \mathbb{R}^d$, with the L_2 norm defined as

$$\|f\| := \left(\sum_{q=1}^{m} \int_D |f_q(x)|^2 dx \right)^{1/2} < \infty. \tag{1}$$

The inner product of $\phi = [\phi_1, \phi_2, \ldots, \phi_m] : \mathbb{R}^d \to \mathbb{R}^m$ and f is defined as

$$\langle f, \phi \rangle := \sum_{q=1}^{m} \langle f_q, \phi_q \rangle = \sum_{q=1}^{m} \int_D f_q(x)\phi_q(x)dx. \tag{2}$$

In the special case that $m = 1$, for a real-valued function $\psi : \mathbb{R}^d \to \mathbb{R}$ defined on $D \subset \mathbb{R}^d$, its L_2 norm becomes $\|\psi\| := (\int_D |\psi(x)|^2 dx)^{1/2}$, while the inner product of ψ_1 and ψ_2 becomes $\langle \psi_1, \psi_2 \rangle = \int_D \psi_1(x)\psi_2(x)dx$.

Given a target function $f : \mathbb{R}^d \to \mathbb{R}^m$, suppose that we have already built a single layer feed-forward network (SLFN) with $L-1$ hidden nodes, i.e, $f_{L-1}(x) = \sum_{j=1}^{L-1} \beta_j g_j(w_j^{\mathrm{T}} x + b_j)$ ($L = 1, 2, \ldots,$ $f_0 = 0$), $\beta_j = [\beta_{j,1}, \ldots, \beta_{j,m}]^{\mathrm{T}}$, and the current residual error, denoted as $e_{L-1} = f - f_{L-1} = [e_{L-1,1}, \ldots, e_{L-1,m}]$, does not reach an acceptable tolerance level. The framework of SCNs provides an effective solution for how to add β_L, g_L (w_L and b_L) leading to $f_L = f_{L-1} + \beta_L g_L$ until the residual error $e_L = f - f_L$ falls into an expected tolerance ϵ. The following Theorem 1 restates the universal approximation property of SCNs (corresponding to Theorem 7 in [21]).

Theorem 1. *Suppose that span(Γ) is dense in L_2 space and $\forall g \in \Gamma$, $0 < \|g\| < b_g$ for some $b_g \in \mathbb{R}^+$. Given $0 < r < 1$ and a nonnegative real number sequence $\{\mu_L\}$ with $\lim_{L\to+\infty} \mu_L = 0$ and $\mu_L \le (1-r)$. For $L = 1, 2, \ldots,$ denoted by*

$$\delta_L = \sum_{q=1}^{m} \delta_{L,q}, \quad \delta_{L,q} = (1 - r - \mu_L)\|e_{L-1,q}\|^2, \quad q = 1, 2, \ldots, m. \tag{3}$$

If the random basis function g_L is generated with the following constraints:

$$\langle e_{L-1,q}, g_L \rangle^2 \ge b_g^2 \delta_{L,q}, \quad q = 1, 2, \ldots, m, \tag{4}$$

and the output weights are evaluated by

Download English Version:

https://daneshyari.com/en/article/4944288

Download Persian Version:

https://daneshyari.com/article/4944288

Daneshyari.com