Information Sciences 473 (2019) 73–86

Contents lists available at ScienceDirect

Information Sciences

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

Robust stochastic configuration networks with maximum correntropy criterion for uncertain data regression

Ming Li[a,c], Changqin Huang[a,b], Dianhui Wang[c,d,*]

[a] School of Information Technology in Education, South China Normal University, Guangzhou, China
[b] Guangdong Engineering Research Center for Smart Learning, South China Normal University, Guangzhou, China
[c] Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia
[d] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Liaoning, China

ARTICLE INFO

Article history:
Received 4 March 2018
Revised 30 August 2018
Accepted 16 September 2018
Available online 19 September 2018

Keywords:
Stochastic configuration networks
Robust data regression
Randomized algorithms
Maximum correntropy criterion
Alternating optimization

ABSTRACT

This paper develops a robust stochastic configuration network (RSCN) framework to cope with data modelling problems when the given samples contain noises or outliers. Technically, RSCNs are built by generalizing the objective function used in our original stochastic configuration networks with maximum correntropy criterion (MCC) induced losses (the proposed algorithm is termed as RSC-MCC). The half-quadratic (HQ) technique is employed to optimize the penalty weights for each training sample, aiming to weaken the impacts caused by the noisy data or outliers throughout the training session. Alternating optimization (AO) methodology is used to renew the RSCN model in company with updated penalty weights determined by HQ methods. The performance of RSC-MCC algorithm is compared with some existing methods, such as the probabilistic robust learning algorithm for neural networks with random weights (PRNNRW), RVFL networks, improved RVFL networks (Imp-RVFL), and our recent work RSCNs with kernel density estimation (RSC-KDE), on two synthetic function approximation examples, four benchmark datasets and one educational data modelling case study (for student learning performance prediction). The experimental results show that RSC-MCC performs more favourably in robust data analytics, and further indicate that our proposed RSCN framework (both RSC-KDE and RSC-MCC) has a good potential for real-world applications.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Although plenty of mathematical theorems guarantee that neural networks can approximate arbitrary functions well, it can be difficult to build a learner model with a sound generalization performance when the given dataset exhibits some uncertainty or ambiguity, for example, data samples are corrupted with noises or outliers [13,17]. Such kind of problem has been extensively investigated in the area of robust statistics and data analytics [4,5,9,10,21,25]. Technically, a lot of efforts are made in developing/improving the cost function to alleviate the negative impacts caused by outliers on building a functional model. These methods include M-estimator and Hampels hyperbolic tangent estimates [4], mean log squared error (MLSE) criterion [21], least median square (LMS) [25], random sample consensus [9,10], and robust support vector regression [5]. An alternative methodology originates from the signal processing area, where in particular, information theoretic

* Corresponding author at: Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia.
E-mail addresses: cqhuang@scnu.edu.cn (C. Huang), dh.wang@latrobe.edu.au (D. Wang).

https://doi.org/10.1016/j.ins.2018.09.026
0020-0255/© 2018 Elsevier Inc. All rights reserved.

learning (ITL) [24] framework is widely used to preserve the nonparametric nature of correlation learning and MSE adaptation. Theoretically, the cost function is still directly estimated from data via a Parzen kernel estimator [22], but it extracts more information from the data for adaptation, and therefore, solutions that are more accurate than MSE in non-Gaussian and nonlinear signal processing. On the whole, both robust statistics and ITL framework lend strong technical supports for robust data modelling in machine learning.

Randomized techniques for training neural networks have been developed in the last decades [2,18,19,23], which can speed up the learning course because of the random assignment of hidden parameters (input weights and biases). Although such kind of method (or model) has been presented on the basis of various viewpoints and frameworks [26], it still has plenty of shortcomings and even some pitfalls in practical implementation [20]. One key bottleneck is how to assign the random parameters (e.g, from which domain or distribution) and how to set the network architecture so that the randomized learner model can have sound learning and generalization capability. For addressing these issues, our recent work developed an advanced randomized learner model, termed as stochastic configuration networks (SCNs), which are built incrementally by assigning the random weights and biases from a support range based on a supervisory mechanism [30]. Later in [31], we investigated the robust data regression problem by considering a weighted least squares problem to evaluate the output weights of SCNs, and successfully developed robust SCNs (RSCNs) with kernel density estimation (KDE) for updating the penalty weights and alternating optimization (AO) for renewing the RSCN model.

A new similarity measure named as correntropy has been widely employed in machine learning and signal processing community [15,22]. In particular, authors in [14] used a maximum correntropy criterion (MCC) for robust face recognition, and empirically verified that MCC has good potential in pattern recognition with training samples contaminated by nonzero mean, non-Gaussian noises, or large outliers. Liu et al. in [22] demonstrated the superiority of MCC compared with mean squared error (MSE) and minimum error entropy (MEE) on uncertain data modelling tasks. In [11], the convergence property of the robust model obtained with MCC was studied under some assumptions, and the role of the scale parameter (in the Gaussian kernel function) in balancing the convergence rate and the model's robustness was investigated. This paper aims at developing an alternative version of RSCNs by using MCC-based objective function in the construction process of SCNs. MCC is usually expressed in the form of Gaussian kernel function, which makes it difficult to analytically evaluate the output weights such as in our original SCNs framework. Technically, we utilize the half-quadratic (HQ) technique [34] for problem-solving, that is, refining the Gaussian-based optimization problem as a weighted linear least squares problem. That allows the output weights to be computed directly based on Moore–Penrose generalized inverse. Then, similar to the tricks we have employed in [31], AO technique is used for iteratively renewing the RSCN model along with penalty weights updating. The performance of our proposed RSC-MCC algorithm is evaluated on two synthetic function approximation examples, four benchmark datasets, and a real-world case study for student learning performance prediction, in comparison with four existing randomized approaches: the probabilistic robust learning algorithm for neural networks with random weights (PRNNRW) [3], RVFL networks [18,23], improved RVFL networks (Imp-RVFL) [8], and our previous work RSCNs with kernel density estimation (RSC-KDE) [31].

The remainder of this paper is organized as follows: Section 2 briefly reviews some technical supports, including our recent work of stochastic configuration networks and some basics on maximum correntropy criterion (MCC). Section 3 details the framework of robust SCNs with the proposal of RSC-MCC algorithm. The effectiveness and advantages of RSC-MCC are demonstrated on some regression problems on low- and high-dimensional synthetic and real datasets in Section 4. Section 5 concludes this work with further remarks.

2. Technical supports

In this section, we review the theoretical details of SCNs and some preliminaries for the similarity measure of maximum correntropy criterion (MCC). The half-quadratic (HQ) technique for solving the MCC induced optimization problem is revisited at the end.

2.1. Revisit of SCNs

Let $\Gamma := \{g_1, g_2, g_3 \ldots\}$ represent a set of real-valued functions, and span(Γ) stands for the associated function space spanned by Γ. $L_2(K)$ 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 $K \subset \mathbb{R}^d$, with the L_2 norm defined as

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

Given another vector-valued function $\phi = [\phi_1, \phi_2, \ldots, \phi_m] : \mathbb{R}^d \to \mathbb{R}^m$, the inner product of ϕ and f is defined as

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

Note that this definition becomes the trivial case when $m = 1$, corresponding to a real-valued function defined on a compact set.

Download English Version:

https://daneshyari.com/en/article/10225716

Download Persian Version:

https://daneshyari.com/article/10225716

Daneshyari.com