Information Sciences 264 (2014) 104–117

Contents lists available at ScienceDirect

Information Sciences

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

Fast decorrelated neural network ensembles with random weights

Monther Alhamdoosh, Dianhui Wang *

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

ARTICLE INFO

Article history:
Received 6 November 2013
Received in revised form 27 November 2013
Accepted 20 December 2013
Available online 31 December 2013

Keywords:
Negative correlation learning
Neural network ensembles
Random vector function link networks
Data regression

ABSTRACT

Negative correlation learning (NCL) aims to produce ensembles with sound generalization capability through controlling the disagreement among base learners' outputs. Such a learning scheme is usually implemented by using feed-forward neural networks with error back-propagation algorithms (BPNNs). However, it suffers from slow convergence, local minima problem and model uncertainties caused by the initial weights and the setting of learning parameters. To achieve a better solution, this paper employs the random vector functional link (RVFL) networks as base components, and incorporates with the NCL strategy for building neural network ensembles. The basis functions of the base models are generated randomly and the parameters of the RVFL networks can be determined by solving a linear equation system. An analytical solution is derived for these parameters, where a cost function defined for NCL and the well-known least squares method are used. To examine the merits of our proposed algorithm, a comparative study is carried out with nine benchmark datasets. Results indicate that our approach outperforms other ensembling techniques on the testing datasets in terms of both effectiveness and efficiency.

1. Introduction

In the last two decades, ensemble learning framework has received considerable attention and resulted in many novel machine learning techniques, for instance, bagging, boosting and random forests [1–4]. The base models of an ensemble can be trained individually or collectively. Ensemble methods also differ in the way that they handle the training data during the learning phase. For example, simple averaging ensembles [5,6] use whole training data to learn base models, but with different parameter settings for each, while bagging, boosting and random forests use different parts of the training dataset for each base model to gain more diversity among ensemble components. Bagging randomly resamples replicates from a given training dataset [1]; while Boosting sequentially resamples the training dataset based on mis-classification probability distribution calculated from previous learnt models [2,3]. Random forests, however, create the training data variability based on the input features rather than the training examples. It selects the splitting features for each node in its base decision trees from a random set of input features [4]. On the other hand, the merging weights of base models can be set equally [7], or learnt by using heuristic means or minimizing a cost function [8,9]. Note that boosting ensembles adapt their averaging weights during the course of training weak learners [3]. It is worthy mentioned that ensemble learning is essential to learn complex domain problems that have complex nonlinear relationships among input variables and output variables. In such cases, the hypotheses space of base models is very large and one single hypothesis cannot model the underlying data distribution.

* Corresponding author. Tel.: +61 3 9479 3034; fax: +61 3 9479 3060.
E-mail addresses: mal-hamdoosh@students.latrobe.edu.au (M. Alhamdoosh), dh.wang@latrobe.edu.au (D. Wang).

0020-0255/$ - see front matter Crown Copyright © 2013 Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.12.016

An important issue in ensemble learning algorithms is how to maintain no duplication of base models, i.e., modeling different regions of the features space by different base models without deteriorating their individual accuracies. Meanwhile, the overall performance of ensemble should be maintained [7]. This is called ensemble diversity in literature [7,10]. Evolutionary approaches are widely used to select ensemble component networks with maximum disagreement among their outputs [8,11–13]. However, such evolutionary-based approaches take long time to converge. In [7] the ensemble base models were trained by using cross-validation datasets and a diversity metric was introduced to trade-off the bias-variance–covariance. The bias-variance–covariance decomposition raised the attention to find ensemble models with minimum covariance among the base models so that the individual model performance can be retained [10].

Negative correlation learning (NCL) [14] amends the cost function with a penalty term that weakens the relationship with other individuals and controls the trade-off among the bias, variance and covariance in the ensemble learning [10]. Later, this idea was extended in [15]. It has been noticed that the proposed approach in [13] can automatically determine an optimal size of ensemble by thoroughly exploring the ensemble hypotheses space using NCL. Recently, a regularized version of the negative correlation learning techniques was proposed in [16], aiming at reducing the overfitting risk for noisy data.

In addition to the diversity concern in ensemble learning, the way of generating the component models is crucial. Neural networks with back-propagation learning algorithms are mostly employed for this purpose. Usually, single-layer feed-forward (SLFN) networks are sufficient to problem solving due to its universal approximation capability [17–19]. Unfortunately, the gradient-based learning algorithms for training SLFNs suffer from local minima problem, slow convergence and very poor sensitivity to learning rate setting. To overcome these difficulties, random vector functional-link (RVFL) networks were proposed [20–22], where the weights between the input layer and the hidden layer can be randomly assigned and no need to be tuned. Then, the well-known least square methods can be used to calculate the output weights [23]. This flat-net architecture universally approximates any continuous function and dramatically reduces the training time [22,24]. Our recent work reported in [12] shows that RVFL-based ensembles take some advantages comparing against other existing ensembling methods. It is interesting to see that in most of the cases the resulting ensemble is composed of few component models (4–12). Note that this finding is obtained by selecting the best RVFL candidates from a pool. Because the best candidates selection is implemented by using genetic algorithms, it is a time consuming process to achieve the final ensemble model. Indeed, this motivates us to do further studies in this direction.

Based on our previous studies, this paper aims to develop a fast solution on building neural network ensembles. Our proposed algorithm (termed as DNNE) randomly initializes the hidden layer parameters of base RVFL networks, and then employs the least square method with negative correlation learning scheme to analytically calculate the output weights of these base networks. A minimum norm least square solution is derived and formulated in a matrix form for computational exercises. A comparative study on data regression is carried out. Results over the testing datasets are promising, and supporting a positive statement on performance assessment among bagging, boosting, simple ensembles and random forests.

The remainder of this paper is organized as follows. Basics on RVFL networks, ensemble learners and the negative correlation learning are reviewed in Section 2. Our solution with a detailed description on the proposed learning algorithm is given in Section 3. To evaluate the performance of our approach, some benchmark datasets are employed in this study and results with comparisons and discussions are presented in Section 4. Finally, we conclude this work in the last section.

2. Background

Learning from data is a process of finding a hypothesis $f(x;\theta)$ that estimates an unknown target function $\psi(x)$, where θ is the model parameters to be tuned.The learnt hypothesis can be single or composite model (i.e., ensemble of base hypotheses). Only few works reported the use of random weights in ensemble base networks [10,20] while it has been widely investigated in single neural networks [21,22,25,26]. The theoretical foundation of randomness in neural networks can be deeply understood from the function approximation task with Monte-Carlo (MC) methods. It has been shown in [22] that any continuous function defined on a compact set can be represented by a limit-integral of multivariate continuous function with integration in the parameters space. MC method approximates this multiple integral by drawing random samples of the parameter vector from uniform distribution defined over the limit-integral domain. More likely, the Monte-Carlo estimated accuracy proportionally improves along with an increase of the number of random samples and the approximation error tends to zero as this number goes to infinity. A special case of this general approximation theory becomes the random vector functional link (RVFL) networks, which can be represented as single-layer feed-forward networks (SLFNs). The multivariate continuous functions in the MC method play a role as the activation functions of hidden neurons in the SLFN, and the input weights and the hidden layer biases of the SLFN correspond to the sampled parameters vector in the MC method. Eventually, the output weights of the SLFN function as the estimated parameters in the MC method.

2.1. Review of random basis function approximators

It has been proved that RVFL networks are universal approximators for continuous functions on compact sets and its approximation error converges to zero with order $O\left(C/\sqrt{L}\right)$, where L is the number of basis functions (hidden neurons) and C is a constant [22,26]. An RVFL network can be defined as a SLFN model,