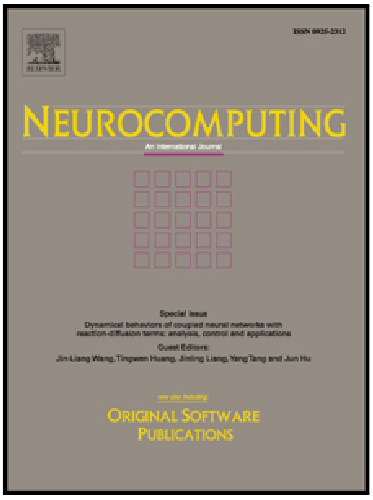
Communicated by Dr. K. Li

Accepted Manuscript



Least Squares Kernel Ensemble Regression in Reproducing Kernel Hilbert Space

Xiang-Jun Shen, Yong Dong, Jian-Ping Gou, Yong-Zhao Zhan, Jianping Fan

PII: S0925-2312(18)30650-7
DOI: 10.1016/j.neucom.2018.05.065
Reference: NEUCOM 19626

To appear in: *Neurocomputing*

Received date: 21 September 2017
Revised date: 24 January 2018
Accepted date: 23 May 2018

Please cite this article as: Xiang-Jun Shen, Yong Dong, Jian-Ping Gou, Yong-Zhao Zhan, Jianping Fan, Least Squares Kernel Ensemble Regression in Reproducing Kernel Hilbert Space, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.05.065

Least Squares Kernel Ensemble Regression in Reproducing Kernel Hilbert Space

Xiang-Jun Shen[a,*], Yong Dong[a], Jian-Ping Gou,[a], Yong-Zhao Zhan[a], Jianping Fan[b]

[a] *School of Computer Science and Communication Engineering, JiangSu University, JiangSu, 212013 China*
[b] *Department of Computer Science, University of North Carolina at Charlotte, NC 28223, USA*

Abstract

Ensemble regression method shows better performance than single regression since ensemble regression method can combine several single regression methods together to improve accuracy and stability of a single regressor. In this paper, we propose a novel kernel ensemble regression method by minimizing total least square loss in multiple Reproducing Kernel Hilbert Spaces (RKHSs). Base kernel regressors are co-optimized and weighted to form an ensemble regressor. In this way, the problem of finding suitable kernel types and their parameters in base kernel regressor is solved in the ensemble regression framework. Experimental results on several datasets, such as artificial datasets, UCI regression and classification datasets, show that our proposed approach achieves the lowest regression loss among comparative regression methods such as ridge regression, support vector regression (SVR), gradient boosting, decision tree regression and random forest.

Keywords:
least squares method, ensemble regression, kernel regression

1. Introduction

In many real-world applications, it is important to predict the value of one feature depending on other measured features [1]. Regression is one of the most fundamental statistical techniques to solve such problems [2, 3], which helps to explore the relationship between inputs and outputs from example data in continuous space. There are many methods [4] which use different strategies to carry out the regression process. These strategies are mainly divided into two categories: single regression models and ensemble regression models [5, 6].

Single regression models can also be categorized into non-kernel and kernel methods. Representative methods in non-kernel methods are ridge regression and lasso regression. For example, Pan et al. [7] proposed an approach based on ridge regression for image reconstruction in computed tomography. Liu et al. [8] extracted plant characteristic gene set, based on lasso logistic regression. Meanwhile kernel methods, such as, kernel ridge regression and support vector regression (SVR), are widely used for their theoretical or experimental results. For example, Burnaev et al. [9] provided a detailed description of a computationally efficient conformal procedure for kernel ridge regression. Zhang et al. [10] presented a new online SVR method called online Laplacian-regularized SVR (online LapSVR).

While in the second broad category of ensemble regression models, the base regression models are combined together for improving accuracy and stability of a single regressor. This has achieved success in many real-world applications, such as random forest regression, gradient boosting regression and decision tree regression. For example, Xu et al. [11] proposed a prediction model based on random forest to analyze some readily available indicator effects on diabetes. Jiang et al. [12] implemented a gradient boosting tree system in the production cluster of Tencent Inc. Wu et al. [13] investigated the nonlinear relationship between land surface temperature and vegetation abundance with the use of decision tree regression approach.

Generally, in single regression methods, the kernel regression methods have better performance than non-kernel regression methods, because of their nonlinear usage of the Reproducing Kernel Hilbert Space (RKHS) [14, 15]. Yukawa et al. [16] proposed adaptive multiple RKHSs learning algorithm by applying Cartesian product. Lv et al. [17] presented a new RKHS sparsity-smoothness penalty with nonlinear function cases. Mitra et al. [18] proposed a novel finite dictionary technique in the RKHS. However, the selection of parameters has great influence on the performance of a single kernel regression method. Therefore, it is a key problem of kernel regression methods in considering selecting suitable kernel types and their parameters.

Considering that ensemble learning methods can improve the performance of a single model, we propose a new learning algorithm that takes advantage of both the ensemble method and kernel learning method. In our proposed kernel ensemble regression method, base kernel regressors are obtained by varying kernel types and their parameters. Ensemble regressor is therefore obtained by combining base kernel regressors. The total least square loss is then minimized in multiple RKHSs. In this way base kernel regressors are finally co-optimized and weighted to build an ensemble regressor. Applying this idea, the problem of finding suitable kernel types and their parameters in base kernel regressors is solved in such an ensemble

*Corresponding author
Email address: xjshen@ujs.edu.cn (Xiang-Jun Shen)