Computational Statistics and Data Analysis 56 (2012) 1681–1691

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Semi-supervised wavelet shrinkage

Kichun Lee [a,*], Brani Vidakovic [b]

[a] *Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, United States*

[b] *Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, United States*

ARTICLE INFO

Article history:
Received 3 February 2011
Received in revised form 15 August 2011
Accepted 10 October 2011
Available online 18 October 2011

Keywords:
K-NN
Manifold-regularization
Semi-supervised learning
Wavelet shrinkage
Denoising

ABSTRACT

To estimate a possibly multivariate regression function g under the general regression setup, $y = g + \epsilon$, one can use wavelet thresholding as an alternative to conventional nonparametric regression methods. Wavelet thresholding is a simple operation in the wavelet domain that selects a subset of coefficients corresponding to an estimator of g when back-transformed. We propose an enhancement to wavelet thresholding by selecting a subset in a semi-supervised fashion in which the neighboring structure and classification function appropriate for wavelet domains are utilized. Wavelet coefficients are classified into two types: *labeled*, which have either strong or weak magnitudes, and *unlabeled*, which have in-between magnitudes. Both are connected to neighboring coefficients and belong to a low-dimensional manifold within the set of all wavelet coefficients. The decision to include a coefficient in the model depends not only on its magnitude but also on the labeled and the unlabeled coefficients from its neighborhood. We discuss the theoretical properties of the method and demonstrate its performance in simulated examples.

1. Introduction

In recent years, both theoretical and applied statisticians have focused on statistical wavelet modeling in estimating true signals from their noisy realizations. The most important property of wavelets is their adaptive locality in both time and frequency, which helps in dealing with the phenomena that change rapidly in both domains. A simple operation in the wavelet domain that selects the subset of wavelet coefficients, which are usually statistically processed, is referred to as wavelet thresholding, or wavelet shrinkage. Shrinkage in the wavelet domain, as a simple, yet powerful tool in nonparametric statistical modeling, utilizes the fact that wavelet transforms are energy-compressing. In other words, most of the true signal variance is described by only a few wavelet coefficients. Previous studies have suggested many approaches to wavelet shrinkage: shrinkage by unbiased estimator of the risk, multiple hypothesis testing, cross-validation techniques, Bayes' rules, to name just a few (Antoniadis et al., 2001). Bayesian approaches (Chipman et al., 1997; Vidakovic and Ruggeri, 2001; Figueiredo and Nowak, 2001) and minimax approaches (Donoho and Johnstone, 1998) among others are suggested for accurate estimation of the true signal. Almost all methodologies involve an underlying statistical model on the wavelet coefficients, and the shrinkage rule represents an optimal action in the adopted statistical paradigm. Many wavelet shrinkage methods based on these approaches are proper thresholding rules, meaning that the inclusion of a wavelet coefficient in the model takes place if its magnitude exceeds a particular threshold. Perturbations on the threshold level always affect the selection of wavelet coefficients, so proper strategies are needed to ensure that the model is neither over- nor under-fitting.

This paper is novel in its fusion of wavelet shrinkage and semi-supervised learning. The resulting shrinkage policy will be termed as semi-supervised wavelet shrinkage. For the coefficients whose magnitudes are close to the adopted threshold, we seek additional information to decide whether they are going to be retained in the model or not. We place

* Corresponding author. Tel.: +1 404 357 5539.
E-mail addresses: skylee1020@gmail.com (K. Lee), brani@bme.gatech.edu (B. Vidakovic).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2011.10.010

this task within the framework of statistical learning and introduce labeled and unlabeled wavelet coefficients. For labeled coefficients, membership in the model is determined while for the unlabeled coefficients, the decision is not clear, so additional information is needed. The unlabeled coefficients are processed under the semi-supervised learning paradigm that incorporates information from their respective neighboring coefficients.

Semi-supervised learning has become very popular in the area of machine learning. It comprises a wide range of methods aimed at enhancing learning from both labeled and unlabeled data and providing better inference (usually in tasks of classification and clustering). While labeled data can be expensive or time-consuming to obtain, unlabeled data are usually easy to collect and may carry information useful for inference, such as in the development of classifiers. In semi-supervised learning, information contained in unlabeled data can be incorporated using a variety of techniques including the expectation maximization (EM) algorithm, transductive support vector machines (SVMs), graph regularization, and others (Zhu, 2005).

In the wavelet context, labeled and unlabeled coefficients are explained using a shrinkage rule with two possible thresholds λ_1 and λ_2, with $\lambda_1 \leq \lambda_2$. The labeled coefficients have two labels, 0 or 1, depending on whether they are excluded or included in the model. Exclusion and inclusion are decided by a thresholding rule using threshold λ_1 and λ_2, respectively. Specifically, the coefficients whose magnitude is less than λ_1 carry the label 0, while the coefficients with magnitude exceeding threshold λ_2 carry the label 1. The labels of wavelet coefficients with a magnitude between λ_1 and λ_2 remain unassigned or undetermined. Such coefficients will be treated as unlabeled.

By taking the approach of manifold regularization, the classification of the unlabeled coefficients is based on the neighborhood content within the manifold structure. We demonstrate that semi-supervised shrinkage (SS) based on background shrinkage with two different thresholds, λ_1 and λ_2, preserves the optimal properties of their background shrinkage while optimizing the performance.

This paper is organized as follows: Section 2 provides short overviews of wavelet shrinkage and the manifold-regularization approach in semi-supervised learning. Section 3 explores the SS rule from the geometric point of view and analyzes asymptotic properties in the context of background thresholding methods. Section 4 provides a discussion on the selection of parameters and comprehensive simulations and comparisons. Section 5 summarizes the results and delineates possible directions for future research. The appendices expand on the technical points in proofs of theorems.

2. Preliminaries

In this section, we present the technical background for wavelet shrinkage and semi-supervised learning needed for proposing semi-supervised shrinkage.

2.1. Wavelet shrinkage

Wavelet shrinkage is motivated by the interplay between the smoothness of the function and magnitudes of its wavelet coefficients. In fact, wavelets are unconditional bases for a range of smoothness spaces, which in plain words means that all information about the smoothness of the function is contained in the rate of decay of magnitudes of its wavelet coefficients. If wavelet transform $\mathcal{W}$ is applied to a vector of noisy measurements $\mathbf{y} = \mathbf{g} + \boldsymbol{\epsilon}$ with normal noise $\boldsymbol{\epsilon}$, the transformed noise $\mathcal{W}\boldsymbol{\epsilon}$ is normal, as well. For the discrete inputs $\mathbf{y}$ of length N, this linear and orthogonal transformation can be fully described by an orthogonal matrix $\mathcal{W}$ of dimension $N \times N$. With $\mathbf{d} = \mathcal{W}\mathbf{y}$ and $\boldsymbol{\theta} = \mathcal{W}\mathbf{g}$, the noise model in the time domain can be reformulated as $\mathbf{d} = \boldsymbol{\theta} + \boldsymbol{\epsilon}$ in the wavelet domain. Wavelet shrinkage methodology, now widely utilized in nonparametric function estimation, estimates $\boldsymbol{\theta}$ from the noisy observations $\mathbf{d}$. The simplest nonlinear wavelet shrinkage technique is thresholding. The two most common thresholding policies are *hard* and *soft* with corresponding rules given by:

$$\theta^{hard}(d, \lambda) = d\mathbf{1}(|d| > \lambda),$$
$$\theta^{soft}(d, \lambda) = (d - \mathrm{sgn}(d)\lambda)\mathbf{1}(|d| > \lambda),$$

where $\mathbf{1}(A)$ is an indicator of relation A.

These two shrinkage mechanisms with properly selected thresholds are characterized by exceptional statistical properties especially under i.i.d. Gaussian noise model (Donoho and Johnstone, 1994). Under the correlated Gaussian noise model, they also have near-optimal behavior in a wide range of function classes (Johnstone and Silverman, 1997). The wavelet shrinkage based on semi-supervised learning, which is developed in this paper, builds on existing shrinkage methods but aims to exploit the hierarchical dependence structure of wavelet coefficients for more accurate model selection.

2.2. Semi-supervised learning

In the real world, we encounter both labeled and unlabeled observations. When we estimate the true link between a label (response) and attributes (variables), it is beneficial to incorporate the unlabeled observations if the labeled and unlabeled attributes come from the same population. This requires a formal model that coherently handles both types of attributes.

In this section, a *Laplacian* kernel, associated with *manifold regularization* (Belkin and Niyogi, 2004), was used to incorporate the information contained in the unlabeled observations. This information refers to low-dimensional manifolds