Computer Vision and Image Understanding 117 (2013) 418–428

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

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

Multi-resolutive sparse approximations of d-dimensional data

Giuseppe Patanè

Consiglio Nazionale delle Ricerche, Istituto di Matematica Applicata e Tecnologie Informatiche, Genova, Italy

ARTICLE INFO

Article history:
Received 2 November 2011
Accepted 3 October 2012
Available online 8 December 2012

Keywords:
Sparse approximation
Support Vector Machine
Image analysis
Least-squares approximation
Reproducing Kernel Hilbert Space
Radial basis functions
Spectral graph theory
Manifold learning

ABSTRACT

This paper proposes an iterative computation of sparse representations of functions defined on $\mathbb{R}^d$, which exploits a formulation of the sparsification problem equivalent to Support Vector Machine and based on Tikhonov regularization. Through this equivalent formulation, the sparsification reduces to an approximation problem with a Tikhonov regularizer, which selects the null coefficients of the resulting approximation. The proposed multi-resolutive sparsification achieves a different resolution in the approximation of the input data through a hierarchy of nested approximation spaces. The idea behind our approach is to combine a smooth and strictly convex approximation of the l_1-norm with Tikhonov regularization and iterative solvers of linear/non-linear equations. Firstly, the iterative sparsification scheme is introduced in a Reproducing Kernel Hilbert Space with respect to its native norm. Then, the sparsification is generalized to arbitrary function spaces using the least-squares norm and radial basis functions. Finally, the discrete sparsification is derived using the eigendecomposition and the spectral properties of sparse matrices; in this case, the computational cost is $O(n\log n)$, with n number of input points. Assuming that the data is supported on a $(d-1)$-dimensional manifold, we derive a variant of the sparsification scheme that guarantees the smoothness of the solution in the ambient and intrinsic space by using spectral graph theory and manifold learning techniques. Finally, we discuss the multi-resolutive approximation of d-dimensional data such as signals, images, and 3D shapes.

1. Introduction

Representing a signal as a linear combination of a set of atoms of a given dictionary is used in a wide range of applications, such as approximation, denoising, and compression. Two main elements characterize the final representation: (i) the properties of the atoms such as linear independence, orthogonality, redundancy, signal-awareness and (ii) the sparseness of the linear representation, which is given by the number of non-null coefficients. Defining sparse representations with respect to dictionaries reacher than an orthogonal basis is also fundamental to represent complex data and to adapt this representation to the features of the data itself. For instance, dictionaries of curvelets [8,9] and bandelets [30,46] are tailored to the local geometric regularity of the input signal and the coefficients of the corresponding sparse representations are useful to identify geometric features; e.g., sharp boundaries and edge orientation in images. Furthermore, the computation of sparse representations with respect to a given dictionary can be combined with an update of its atoms in order to improve the data fitting [1]. Main applications of sparse representations in computer vision and image understanding include face recognition [61], data segmentation [20,50], image super-resolution [62], denoising [37], and classification [35,36].

Given a signal $f:\mathbb{R}^d\to\mathbb{R}$ and a dictionary $\mathcal{B}:=\{\varphi_i(\mathbf{x})\}_{i=1}^n$ of atoms, sparse coding refers to the problem of computing the coefficients $\mathbf{a}:=(a_i)_{i=1}^n$ of the function $g(\mathbf{x})=\sum_{i=1}^n a_i\varphi_i(\mathbf{x})$ that approximates f, involves the smallest number of atoms, and provides the highest accuracy among all the approximations of f generated by $\mathcal{B}$. In this context, compressive sampling theory [8,15] has shown that signals can be accurately approximated from a number of samples that is lower than the one imposed by the Nyquist sampling theory.

According to [14,52], the coefficient vector $\mathbf{a}$, which defines the *sparse representation* $g:\mathbb{R}^d\to\mathbb{R}$ of f, solves the minimization problem

$$\arg\min_{\mathbf{a}\in\mathbb{R}^n}\{E(f,g)+\epsilon\|\mathbf{a}\|_0\},\qquad g(\mathbf{x}):=\sum_{i=1}^n a_i\varphi_i(\mathbf{x}),\tag{1}$$

where the term $E(f,g)$ is the approximation error between f and g with respect to the loss function $E(\cdot,\cdot)$; the sparsification order $\|\mathbf{a}\|_0$ is given by the number of non-null coefficients; and the positive constant ϵ controls the trade-off between these two terms.

To measure the approximation error between the maps f and g, common choices are the native distance $E(f,g):=\frac{1}{2}\|f-g\|_{\mathcal{H}}^2$ in a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$; the l_2-norm $E(f,g):=\frac{1}{2}\|\mathbf{f}-\mathbf{g}\|_2^2$ of the values $\mathbf{f}:=(f(\mathbf{x}_i))_{i=1}^n$, $\mathbf{g}:=(g(\mathbf{x}_i))_{i=1}^n$ at the points of $\mathcal{P}:=\{\mathbf{x}_i\}_{i=1}^n$; and the ϵ-insensitive cost function [14,52]

E-mail address: patane@ge.imati.cnr.it

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2012.10.012

$$E(f,g) := \sum_{i=1}^{n} \Gamma(f(\mathbf{x}_i) - g(\mathbf{x}_i)), \quad \Gamma(t) := |t|_\epsilon := \begin{cases} 0 & \text{if } |t| < \epsilon, \\ |t| - \epsilon & \text{otherwise.} \end{cases}$$

Since the minimization of the objective function in Eq. (1) is NP-hard, the sparsification term $\|\mathbf{a}\|_0$ is usually approximated by the l_p-norm $\|\mathbf{a}\|_p := \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p}$ and the corresponding sparsification results in a convex minimization problem. On the one hand, for $0 \leqslant p < 1$ the l_p-norm is not strictly convex [25,47–49] and the corresponding problem has local extrema that might be identified as solutions during the search of the global minimum. On the other hand, the l_1-norm guarantees the uniqueness of the solution and provides a representation sparser than the l_2-norm. To avoid oversampling, an iterative re-weighted l_2-norm minimization, which provides a sparsity percentage lower than the l_2-norm, has been proposed in [28]. From a general perspective, the l_1-norm is preferable to the l_2-norm sparsification term because the former avoids to penalize outliers in the sampled data and to distribute the residual error in the objective functional [8,15,58]. Although the l_0-norm provides the sparsest solution, the assumption of dealing with a bounded noise generally guarantees that the l_1-norm sparse representations are significative and stable to noise and outliers.

The basis pursuit de-noising [11], regularized logistic regression [24,41,45,51], standard [39] and orthogonality matching pursuit methods [10,38,43] use the l_1-norm as sparsification term. Since the l_1-norm is not differentiable at zero, the sparsification is converted to a constrained optimization problem [32], whose number of unknowns is twice the number of input variables. Alternatively, the l_1-norm is approximated by a second order Taylor expansion of the objective function [24], which is minimized using the least-squares angle regression [18] and the quasi-Newton algorithm [2]. The sparsification problem is also solved through an incremental approach [32], which is based on the conjugate gradient and avoids the discontinuity of the first order derivatives of the l_1-norm. Alternative approaches apply the maximum *a posteriori* estimation [33,34,42] and uncertainty criteria [16,17,19,21,26]. Finally, the probabilistic Bayesian learning framework [57] is capable of further increasing the sparsification rate with respect to SVMs and applies to arbitrary kernels.

Aims and contributions This paper discusses an iterative computation of sparse and multi-resolutive representations of an arbitrary function, which achieves a different resolution through a hierarchy of nested approximation spaces. The proposed approach exploits a formulation [22] of the sparse approximation problem equivalent to Support Vector Machine and based on Tikhonov regularization. Through this equivalent formulation, the sparsification reduces to an approximation problem with a Tikhonov regularizer, which selects the null coefficients of the resulting approximation. The idea behind our sparsification is to combine a smooth and strictly convex approximation of the l_1-norm with Tikhonov regularization and iterative solvers of linear or non-linear equations. The proposed approach also guarantees good generalization performances and applies to arbitrary function spaces, whose basis is not necessarily associated to a Mercer kernel. Finally, it provides a sequence of nested approximation spaces, which are generated by those functions selected during the computation of the sparsified solution. We also discuss the multi-resolutive approximation of d-dimensional data such as signals, images, and 3D shapes.

To introduce the sparsification scheme, we firstly assume that $\mathcal{H}$ is a Reproducing Kernel Hilbert Space (RKHS) [3]; in this case, the native norm of $\mathcal{H}$ allows us to enforce the accuracy and smoothness of the sparse approximation. Using the equivalence between Support Vector Machine and Tikhonov regularization [22] in a RKHS, we approximate a real-valued function with sparse linear models, whose coefficients are fitted using a smoothed version of the l_1-regularization. This aim is achieved by replacing the l_1-norm with a smooth and strictly convex approximation; then, the corresponding sparsification functional is exactly evaluated and no approximation is required. Finally, the sparsification problem is converted into a system of non-linear equations, whose sparse coefficient vector is computed by applying a fixed point iteration and solving a sequence of linear systems.

Using radial basis functions and least-squares techniques, the second part of the paper generalizes the iterative sparsification scheme to arbitrary function spaces, which are not necessarily associated to Mercer kernels. Assuming that the data is supported on a $(d-1)$-dimensional manifold, we also derive a variant of the proposed approach that guarantees the smoothness of the solution in the ambient and intrinsic space by using spectral graph theory and manifold learning techniques. Diagonalizing the Gram matrix of the sparsification normal equation, the unknown coefficients become independent; i.e., each non-linear equation involves only one unknown and its solution is computed in explicit form.

Applying iterative solvers instead of decomposition methods for constrained convex minimization problems has the following advantages with respect to previous work. The computational cost of the overall framework is $O(r(n + n\log n))$ instead of $O(n^{3.5})$, where n and r, $r \ll n$, are the number of input data and steps of the iterative sparsification scheme, respectively. The solution of the sparsification system is well-conditioned as a matter of the underlying regularization framework and based on a global sparsification procedure, which avoids time-consuming and *a posteriori* local updates of the model. Furthermore, at each iteration the update of the coefficient matrix involves only its diagonal elements, takes $O(n)$ time, and preserves its sparsity and symmetric structure. Finally, the input variables are not duplicated, thus reducing the memory allocation, which is one of the main drawbacks in case of a large amount of data. Since each iteration provides an approximate reconstruction of the input function $f : \mathbb{R}^d \to \mathbb{R}$, the iterative solver induces a hierarchy of sparse representations $(g^{(r)})_r$ of f, which belong to a sequence of nested spaces $(\mathcal{H}_r)_r$, $\mathcal{H}_{r+1} \subseteq \mathcal{H}_r$, $r \geqslant 1$.

The paper is organized as follows. First, we introduce the proposed sparsification scheme in Reproducing Kernel Hilbert Spaces (Section 2). Then, we derive a least-squares variant and its discrete counterpart (Section 3). Finally, we outline open issues and future work (Section 4).

2. "Iterative" sparse approximation in Reproducing Kernel Hilbert Spaces

Replacing the l_1-norm with a smooth approximation, we define an iterative sparsification scheme (Section 2.1) in a RKHS with respect to its native norm. Then, we discuss the iterative computation of the sparsified solution (Section 2.2), and the multi-resolutive structure of the sparsification scheme (Section 2.3). Finally, the generalization of the sparsification scheme to arbitrary function spaces is addressed in Section 3.

2.1. Sparsification in Reproducing Kernel Hilbert Spaces

Let $\mathcal{H}$ be a Reproducing Kernel Hilbert Space [3] endowed with the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and norm $\| \cdot \|_{\mathcal{H}}$ induced by a positive definite, symmetric kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Common choices of $K(\cdot,\cdot)$ are the Gaussian $K(\mathbf{x},\mathbf{y}) := \exp(-\|\mathbf{x}-\mathbf{y}\|_2^2)$, polynomial $K(\mathbf{x},\mathbf{y}) := (1 - \langle \mathbf{x},\mathbf{y} \rangle_2)^s$, and compactly supported [40,53] kernels. Let $g : \mathbb{R}^d \to \mathbb{R}$

$$g(\mathbf{x}) := \sum_{i=1}^{n} a_i K(\mathbf{x}, \mathbf{x}_i), \qquad \mathbf{a} := (a_i)_{i=1}^n \in \mathbb{R}^n,$$

be a map in the linear space $\mathcal{H}_n \subseteq \mathcal{H}$ generated by the basis $\mathcal{B} := \{\varphi_i(\mathbf{x})\}_{i=1}^n$, where each function $\varphi_i(\mathbf{x}) := K(\mathbf{x},\mathbf{x}_i)$ is induced by