SPARSE CORRELATION KERNEL RECONSTRUCTION

Constantine Papageorgiou, Federico Girosi, Tomaso Poggio

Center for Biological and Computational Learning and Artificial Intelligence Laboratory
MIT
Cambridge, MA 02139
{cpapa,girosi,tp}@ai.mit.edu

ABSTRACT

This paper presents a new paradigm for signal reconstruction and superresolution, Correlation Kernel Analysis (CKA), that is based on the selection of a sparse set of bases from a large dictionary of class-specific basis functions. The basis functions that we use are the correlation functions of the class of signals we are analyzing. To choose the appropriate features from this large dictionary, we use Support Vector Machine (SVM) regression and compare this to traditional Principal Component Analysis (PCA) for the task of signal reconstruction. The testbed we use in this paper is a set of images of pedestrians. Based on the results presented here, we conclude that, when used with a sparse representation technique, the correlation function is an effective kernel for image reconstruction.

1. INTRODUCTION

This paper presents Correlation Kernel Analysis (CKA), a new paradigm for signal reconstruction and compression that is based on the selection of a sparse set of bases from a large dictionary of class-specific basis functions. The concept of sparsity enforces the requirement that, given a certain reconstruction error, we should choose the smallest subset of basis functions that yields a reconstruction with this error. The problem of signal reconstruction is formulated as one where we are given only a small, possibly unevenly sampled, subset of points in a signal where the goal is to accurately reconstruct the entire signal.

The signal approximation problem we present assumes that we have prior information about the class of signals we are reconstructing or compressing in the form of the correlation function of the class of signals to which this signal belongs, as defined by a representative set of signals from this class [9] [10]. For this paper, the signals that we will be looking at are images of pedestrians [6] [4] [8]. Using an initial set of pedestrian images, we compute the correlation function and use the pointwise-defined functions as the dictionary of basis functions from which we can reconstruct subsequent out-of-sample images of pedestrians. Our choice of using the correlation kernel can be motivated from a Bayesian point of view.

To approximate or reconstruct an image, rather than using the entire set of correlation-based basis functions comprising the dictionary we choose a small subset of the kernels via the criteria of sparsity. We obtain a sparse representation by approximating the signal using the Support Vector Machine (SVM) [1] [11] formulation of the regression problem.

The results presented in this paper can be useful in low-bandwidth videoconferencing, image de-noising, reconstruction in the presence of occlusions, signal approximation from sparse data, as well as in superresolving images. This technique is an alternative to traditional means of function approximation and signal reconstruction, such as Principal Components Analysis (PCA), for a wider class of signals than just images.

2. GENERALIZED CORRELATION KERNELS

To reconstruct or compress a function f, we use information about the class of pointwise mean-normalized signals that f is a part of, derived from a set of representative examples from that class. This information is in the form of the correlation function of the signals in the class:

$$R(\mathbf{x}, \mathbf{y}) = E[(f_\alpha(\mathbf{x}) - \mu(\mathbf{x}))(f_\alpha(\mathbf{y}) - \mu(\mathbf{y}))] \qquad (1)$$

where f_α are instances of the class of functions to which f belongs, $\mathbf{x}$ and $\mathbf{y}$ are coordinates in the 2-dimensional signal, and μ are the point means across the class of functions: $\mu(\mathbf{x}) = E[f_\alpha(\mathbf{x})]$.

We can also generate the eigen-decomposition of the symmetric, positive definite correlation matrix by solving

$$\int d\mathbf{x} R(\mathbf{x}, \mathbf{y}) \phi_n(\mathbf{x}) = \lambda_n \phi_n(\mathbf{y}) \qquad (2)$$

where ϕ_n are the eigenvectors and λ_n are the eigenvalues of the system. After generating this decomposition, we can write R in the form,

$$R(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{M} \lambda_n \phi_n(\mathbf{x}) \phi_n(\mathbf{y}) \qquad (3)$$

where $M \leq \infty$.

The set of functions ϕ_n are ordered with decreasing positive eigenvalue λ_n and are normalized to form an orthonormal basis for the correlation function of f_α.

The correlation function R, which is positive definite, induces a Reproducing Kernel Hilbert Space (RKHS) that allows us to approximate the function f as [10]:

$$f(\mathbf{x}) = \sum_{i=1}^{N} c_i R(\mathbf{x}, \mathbf{x}_i) \qquad (4)$$

where i ranges over pixel locations in the image; R is the reproducing kernel in this space and the norm is:

$$\|f\|_R^2 = \sum_{n=1}^{M} \frac{c_n^2}{\lambda_n} \qquad (5)$$

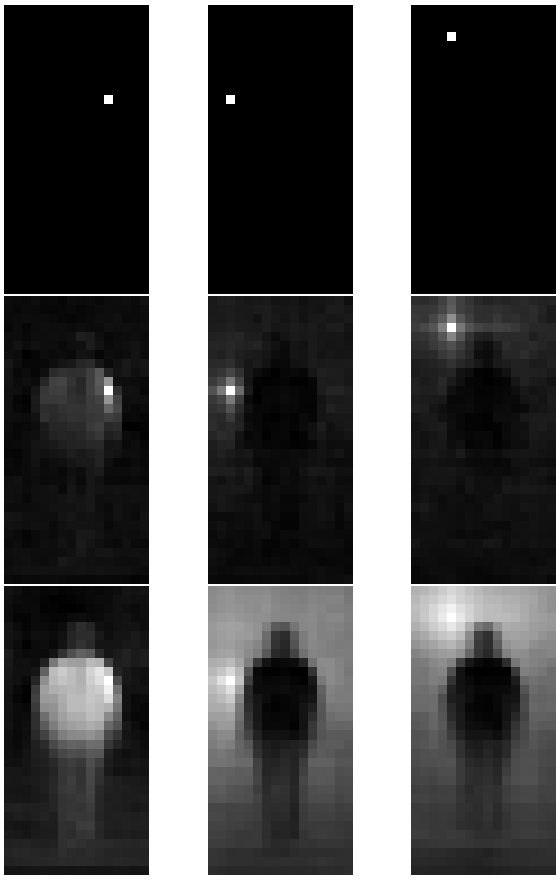


Figure 1: Examples of the correlation kernels we can compute. Each column shows the kernels, $R_d((x_1 = a, x_2 = b), \mathbf{y})$, for a specific (a, b) where $d = 0.0$, $d = 0.5$, and $d = 1.0$ in the top, middle, and bottom rows, respectively. These images demonstrate that $d = 1.0$ corresponds to a very smooth kernel, while $d = 0.0$ is highly localized.

We can obtain a wider class of kernels spanning exactly the same space of functions as the correlation function in Equation 3 by varying the degree of λ_n, which in effect controls the prior information regarding the strength of each eigenfunction, an observation due to [9]. We therefore define the *generalized correlation kernel* as:

$$R_d(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{M} (\lambda_n)^d \phi_n(\mathbf{x}) \phi_n(\mathbf{y}) \tag{6}$$

and notice that the parameter d controls the locality of the kernel; for small d, R_d approaches a delta function in the space of ϕ_n, and as d gets larger, R_d gets smoother[1].

Each of these correlation kernels is a function in four variables (x_1,x_2,y_1,y_2) so, to effectively visualize them, we hold the x_1 and x_2 positions constant and vary y_1 and y_2. Figure 1 shows several examples of the kernels generated with varying d, for a set of 924 grey-level 128×64 images of pedestrians that have been normalized to the same scale and position; this database has been used in [6], [4], and [8]. The progressive delocalization of the kernels when d is varied from 0.0 to 1.0 is evident in these figures.

[1]This particular parameterization is one of many possibilities.

3. BAYESIAN MOTIVATION

Our choice of the correlation function, R, as the kernel can be motivated from a Bayesian perspective; see [12] and [10] for background material. Consider the general regularization problem,

$$\min_{f \in \mathcal{H}} H[f] = \sum_{i=1}^{N} (y_i - f(\mathbf{x}_i))^2 + \gamma \|f\|_K^2 \tag{7}$$

where, in a Bayesian interpretation, the data term is a model of the noise and the stabilizer is a prior on the regression function f. If we assume that the data, y_i, are affected by additive independent gaussian noise, then we can show that the stabilizer measures the Mahalanobis distance of f from the mean signal. This also corresponds to a zero mean multivariate gaussian density on the Hilbert space of functions defined by R and spanned by ϕ_n, e.g., the space spanned by the principal components introduced in Section 2. From a Bayesian point of view, under the assumption of gaussian noise, R is the right kernel to use, whenever it is available. It is important to note that in our SVM and BPDN formulations, we use gaussian priors but do not assume gaussian additive noise in the data.

4. SUPPORT VECTOR MACHINES AND SPARSITY

The operational definition of a sparse representation that we will use in the context of regression is the smallest subset of elements from a large dictionary of features such that a linear superposition of these features can effectively reconstruct the original signal. Here, we present a brief introduction to Support Vector Machine regression; for a more in depth treatment of this subject, the reader is referred to [1], [11], [2], [3].

Given a kernel K that defines a RKHS and with the appropriate choice of the scalar product induced by K, the empirical risk minimization regularization theory framework suggests to minimize the following functional:

$$H[f] = \frac{1}{N} \sum_{i=1}^{N} \| z_i - f(\mathbf{x}_i) \|_{L_2}^2 + \gamma \|f\|_K^2 \tag{8}$$

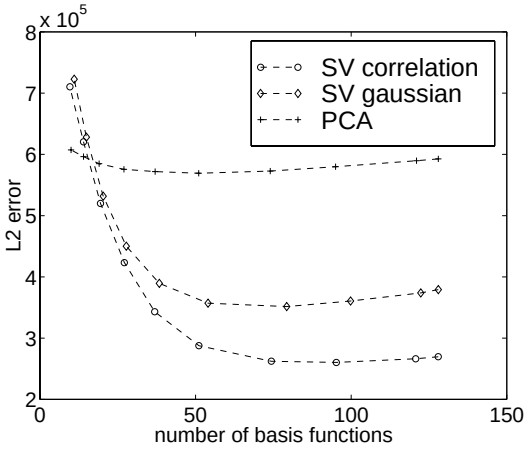
where $\|f\|_K^2$ is as defined in Section 2. This corresponds to minimizing the sum of the empirical error measured in L_2 and a smoothness functional. The Support Vector Machine regression formulation minimizes a similar functional, differing only in the norm on the data term; instead of using the L_2 norm, the following ϵ-insensitive error function, called the L_ϵ norm, is used:

$$|z_i - f(\mathbf{x}_i)|_\epsilon = \begin{cases} 0 & \text{if } |z_i - f(\mathbf{x}_i)| < \epsilon \\ |z_i - f(\mathbf{x}_i)| - \epsilon & otherwise \end{cases} \tag{9}$$

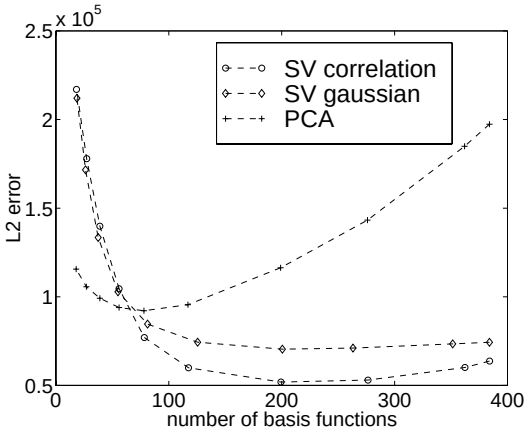
The functional that is minimized is therefore:

$$H[f] = \frac{1}{N} \sum_{i=1}^{N} |z_i - f(\mathbf{x}_i)|_\epsilon + \gamma \|f\|_K^2 \tag{10}$$

yielding a function of the form:



(a)



(b)

Figure 2: Out-of-sample L_2 reconstruction error comparison between SVM with correlation kernel $R_{1.0}$, SVM with gaussian kernel ($\sigma = 3.0$), and PCA, where the input is a random sampling of the original image. Each of these figures represents a different sized sampling, (a) $\frac{1}{4}$ of the image as input and (b) $\frac{3}{4}$ of the image as input.

$$f(\mathbf{x}) = \sum_{i=1}^{N'} c_i R(\mathbf{x}, \mathbf{x}_i) \tag{11}$$

where the coefficients $\mathbf{c}$ are obtained by solving a quadratic programming problem [11] [5] [3]. Depending on the value of the sparsity parameter γ, the number of c_i that differ from zero will be smaller than N; the data points associated with the non-zero coefficients are called *support vectors* and it is these support vectors that comprise our sparse approximation.

5. RECONSTRUCTION

In the case of image reconstruction and compression when we do not assume any prior knowledge (other than that we are considering images), we can use techniques like JPEG, wavelets, and regularization using a spline or gaussian kernel. When we have statistical information on the class of functions we are reconstructing, as in the case of the correlational structure of the class to which the image to be reconstructed belongs, we may be able to obtain better compression by using this information.

The generalized correlation kernels are generated from a training set of 924 grey-level 32×16 images of pedestrians. We test the correlation kernels with $d = 1.0$ by analyzing the SVM reconstruction of pedestrian images not in the training set and comparing to PCA. For each image in the out-of-sample test set, we randomly partition the pixels into a set that has M pixels – the input set, F_{input} – and a set consisting of the remaining $(N - M)$ pixels – the test set, F_{test}.

In the case of the SVM, to find the sparse set of basis functions that minimizes the error over the input subset, F_{input}, we obtain the coefficients of reconstruction by minimizing:

$$H[f] = \frac{1}{M} \sum_{i=1}^{M} |F_{input}(\mathbf{x}_i) - f(\mathbf{x}_i)|_\epsilon + \frac{1}{C} \|f\|_K^2 \tag{12}$$

where,

$$f(\mathbf{x}) = \sum_{i=1}^{M} c_i R(\mathbf{x}, \mathbf{x}_i) \tag{13}$$

The portion of the coefficients, c_i, that will be 0 is determined by the variable C.

Out-of-sample performance in each case is determined by reconstructing the full image and measuring the error over the pixels in F_{test}. We measure performance as the error achieved with respect to the number of basis functions used in the above formulations. In SVM regression, the number of basis functions is varied by changing the ϵ parameter. To compare with PCA-based reconstruction, for a given ϵ, we use, as the number of principal components for the reconstruction, the number of support vectors found in the SVM formulation. In our experiments, the size of the input set is varied as $\frac{1}{4}N$ and $\frac{3}{4}N$; error is measured in L_2. As a benchmark meant to ensure that the performance of the system using SVM with the correlation kernels is not due exclusively to the SVM machinery, we also show the results using SVM with gaussian kernels.

The results of these reconstructions, averaged over 50 out-of-sample images, are shown in Figures 2a and b for the cases of using $\frac{1}{4}$ and $\frac{3}{4}$ of the pixels as input, respectively. From these performance results, we can see that, even though the PCA formulation minimizes L_2 error and SVM regression is minimizing error in the RKHS induced by the ϵ-insensitive norm, SVM performs better than PCA even when measuring error in L_2 on out-of-sample test data. Furthermore, SVM with the correlation kernels outperforms SVM with gaussian kernels, showing that the correlation kernels encode important prior information on the pedestrian class. The difference in performance is most pronounced for the reconstructions that use the smallest input set.

Figure 3 presents an extreme case where the input data is a random set of only $\frac{1}{16}$th (6.25%) of the image pixels; here, a higher resolution image (64×32) is used. The SVM reconstruction with correlation kernels recovers more of the structure of the pedestrian than PCA, due to the smoothness preserving properties of the SVM approach to function approximation [11].

It is possible to use this same framework to *superresolve* an image, that is, reconstruct it at a finer level of detail than was originally present in the image. This could be useful if, for instance, we have an image of a person's face that is too small for us to be able to

(a) (b) (c) (d)

Figure 3: Reconstruction comparison for a higher resolution image (64×32) using identical random sets of $\frac{1}{16}$th of the original pixels as input; (a) the original image, (b) PCA reconstruction with 74 basis functions, (c) SVM reconstruction with 74 basis functions ($\epsilon = 10$ for the SVM), (d) locations of the support vectors are denoted as black values. With a small subset of the original image as input, the SVM reconstruction is clearly superior to the PCA reconstruction.

recognize who it is; after superresolving the image, the details that emerge could allow us to recognize the person. For brevity, we refer the reader to [7] for the details of this work.

6. CONCLUSION

We have shown that the use of class-specific correlation-based kernels, when combined with the notion of sparsity, results in a powerful signal reconstruction technique. In a comparison to a traditional method of signal approximation, Principal Components Analysis, our approach achieves a more sparse representation for a given level of error.

Our approach of using a dictionary of class-specific correlation kernels to obtain sparse representation of a signal leads to an interesting question: could this sparse representation that has been generated to *approximate* a signal be used to *classify* different signals? In other words, is the representation of pedestrians via sparse sets of correlation-based basis functions different enough from the representation of other objects (or all other objects), so that it can be used as a model for that class of objects? The representations we generate are derived through an argument that minimizes error for reconstructing the image. This, however, says nothing about the ability of that same representation to be used to differentiate images of different objects. Whether or not this can be done is an open question; [7] presents a preliminary discussion of this approach.

7. REFERENCES

[1] B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifier. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pages 144–52. ACM, 1992.

[2] C.J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. In Usama Fayyad, editor, *Proceedings of Data Mining and Knowledge Discovery*, pages 1–43, 1998.

[3] F. Girosi. An equivalence between sparse approximation and Support Vector Machines. *Neural Computation*, 10(6):1455–1480, 1998.

[4] M. Oren, C.P. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *Computer Vision and Pattern Recognition*, pages 193–99, 1997.

[5] E. Osuna, R. Freund, and F. Girosi. Support Vector Machines: Training and Applications. A.I. Memo 1602, MIT Artificial Intelligence Laboratory, 1997.

[6] C.P. Papageorgiou. Object and Pattern Detection in Video Sequences. Master's thesis, MIT, 1997.

[7] C.P. Papageorgiou, F. Girosi, and T. Poggio. Sparse Correlation Kernel Analysis and Reconstruction. A.I. Memo 1635, MIT Artificial Intelligence Laboratory, 1998. (CBCL Memo 162).

[8] C.P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Proceedings of International Conference on Computer Vision*, 1998.

[9] P. S. Penev and J. J. Atick. Local Feature Analysis: A general statistical theory for object representation. *Neural Systems*, 7(3):477–500, 1996.

[10] T. Poggio and F. Girosi. A Sparse Representation for Function Approximation. *Neural Computation*, 10(6):1445–1454, 1998.

[11] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer Verlag, 1995.

[12] G. Wahba. *Spline Models for Observational Data*. Series in Applied Mathematics, Vol. 59, SIAM, Philadelphia, 1990.