Wavelet Kernels in RKHS  
Noyaux d’Ondelettes dans les RKHS

Alain Rakotomamonjy, Xavier Mary and Stéphane Canu

Perception, Systèmes et Informations  
INSA de Rouen  
Avenue de l’université  
76801 Saint Etienne du Rouvray France  
alain.rakotomamonjy@insa-rouen.fr

Abstract: Cet article traite de la construction d’espaces de Hilbert à noyaux reproduisants à partir d’opérateurs. Un point de vue théorique ainsi qu’une méthode constructive sont présentés et permettent de proposer des espaces d’hypothèses pour l’apprentissage engendrés par une famille d’ondelettes. Des exemples sur des problèmes-jouets illustrent cette mise en œuvre pratique.

Keywords: Reproducing Kernel, Wavelet, Support Vectors Machines.

1 Introduction

The purpose of this paper is to present hypothesis spaces built from operators in which one looks for the solution of a learning from examples problem. Learning from examples can be viewed as the estimation of the functional dependency between an input \( x \) and an output \( y \) of a system given a set of examples \( \{(x_i, y_i), x_i \in \mathcal{X}, y_i \in \mathcal{Y}, i = 1 \ldots \ell \} \). Besides, we suppose that the examples are independently drawn according to an unknown probability density function. This problem of functional estimation from sparse data is an ill-posed problem and a classical way to turn it in a well-posed one is to use regularization theory. In this context, the solution of the problem is the function \( f \) that minimizes the regularized empirical risk:

\[
R[f] = \frac{1}{\ell} \sum_{i=1}^{\ell} C(y_i, f(y_i)) + \lambda \|f\|_H^2
\]  

(1)

where \( C(\cdot, \cdot) \) is a cost function, \( H \) is a Reproducing Kernel Hilbert Space and \( \lambda \) a regularization parameter. Depending on the cost function used, this minimization problem leads either to SVM or Regularization networks (Evgeniou, Pontil & Poggio
2000). Under general conditions (Kimeldorf & Wahba 1971), the solution of this minimization problem is:

\[ f(x) = \sum_{i=1}^{\ell} a_i K(x, x_i) + \sum_{j=1}^{m} b_j g_j(x) \]  (2)

where \( K(\cdot, \cdot) \) is a semi-positive definite kernel associate to the RKHS \( \mathcal{H} \) and \( \{g_j\}_{j=1}^{m} \) a set of functions spanning the null space of the functional \( \|f\|_\mathcal{H} \). Thus, RKHS plays a central role as it describes the hypothesis space where one looks for the function \( f \). Several papers highlight the importance of using a wisely chosen RKHS as it influences largely the generalization capability of \( f \) (Scholkopf & Smola 2001, Vapnik 1998). Our aim in this paper is to present a way for building RKHS by means of a Carleman operator, thus allowing a practitioner to build a kernel adapted to its problem at hand.

2 Building Kernels from Operators

Let \( B \) be an Hilbert space endowed with inner product \( \langle \cdot, \cdot \rangle_B \) so that

\[ \forall f \in B, \|f\|_B < \infty \]

Let define an indexed family of function \( \Gamma_t(\cdot) \in B \) indexed by \( t \in \mathcal{X} \) and a linear mapping \( T : \)

\[ T : B \rightarrow \mathbb{R}^\mathcal{X} \]

\[ f \rightarrow g(\cdot) \text{ so that } g(t) = T f(t) = \langle \Gamma_t(\cdot), f(\cdot) \rangle_B \]  (3)

We decompose \( B = Ker(T) \oplus \mathcal{M} \) and we call \( S \) the restriction of \( T \) so that:

\[ S : \mathcal{M} \rightarrow Im(T) \]

\[ f \rightarrow g(\cdot) = S f = T f \]  (4)

\( S \) is a bijective operator and we define \( \mathcal{H} \) as \( \mathcal{H} \triangleq Im(T) \). If one endows \( \mathcal{H} \) with the following inner product:

\[ \forall g_1, g_2 \in \mathcal{H} \quad \langle g_1, g_2 \rangle_\mathcal{H} = \langle S f_1, S f_2 \rangle_\mathcal{H} = \langle f_1, f_2 \rangle_B \]  (5)

then \( \mathcal{H} \) is a reproducing kernel Hilbert space with kernel \( K(\cdot) \):

\[ K(t, s) = \langle \Gamma(t, \cdot), \Gamma(s, \cdot) \rangle_\mathcal{H} \]  (6)

A detailed proofs of these propositions can be found in (Mary, Brucq & Canu 2002).
3 Practical Construction of Kernel in $L_2(\mathcal{X})$

Let us restrict to cases where $\mathcal{B} = L_2$ and $\mathcal{H} \subset L_2$. Consider the family $\{\phi_i\}$ as an orthonormal basis of $L_2$. $\mathcal{M}$ is a subspace of $L_2$ and let $\psi_i$ be a basis of $\mathcal{M}$. Recall that $\mathcal{M}$ and $\mathcal{H}$ have the same dimension (due to the bijective property of $S$). As $\Gamma_i(\cdot)$ is a function of $L_2$ and $\mathcal{H} \subset L_2$, one can write:

$$\forall t \in \Omega, \quad \Gamma_i(\cdot) = \sum_{i,j} \alpha_{i,j} \phi_j(t) \phi_i(\cdot)$$

(7)

This equation supposes that $\phi_i(t)$ exists and is well defined for any $t \in \mathcal{X}$. In other words, this means that the considered orthonormal basis must be defined pointwise. Then reproducing kernel of $\mathcal{H}$ becomes:

$$K(s,t) = \langle \Gamma_i(\cdot), \Gamma_s(\cdot) \rangle = \sum_{i,j,n} \alpha_{i,j} \alpha_{i,n} \phi_j(t) \phi_n(s)$$

4 Applications on a toy problems

In the following, we are willing to validate the effectiveness of these way of building kernels on three classical classification problems. We focused only on the practical application of these new hypothesis spaces and we have not provided much efforts for optimising them in order to achieve state-of-art results on the benchmark datasets. The first toy problem is the artificial checkers problem whereas the two latter are the breast cancer and diabetis datasets that are parts of the IDA datasets (Rätsch, Onoda & Müller 2001) and they are already provided with 100 realizations of training and test sets. The induction algorithm for solving the classification problem is Support Vector Machines. For each problem, an hypothesis space associated with a kernel $K$ has been built from an operator $\Gamma_i(\cdot)$ that can be expressed as:

$$\Gamma_i(\cdot) = \sum_{j,k} \psi_{j,k}(t) \psi_{j,k}(\cdot)$$
Table 1: Generalization performance of wavelet kernels and gaussian kernel.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Wavelet Kernel</th>
<th>Gaussian Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checkers</td>
<td>9.01 ± 4.00</td>
<td>15.59 ± 1.6</td>
</tr>
<tr>
<td>Breast Cancer</td>
<td>28.81 ± 4.56</td>
<td>26.00 ± 4.7</td>
</tr>
<tr>
<td>Diabetis</td>
<td>28.04 ± 2.13</td>
<td>23.57 ± 1.7</td>
</tr>
</tbody>
</table>

where $\psi_{j,k}$ is a translated wavelet of resolution $j$. The presented results are obtained from the averaging of 100 trial’s classification error. For the checkers problem, the width of the gaussian kernel has been optimized by a validation procedure whereas for the others two, the best hyperparameters found by Raestch et al have been used (Rätsch et al. 2001).

5 Conclusions and Perspectives

A simple way for constructing reproducing kernel hilbert spaces from operators has been given in this paper and it allows to create kernels spanned by wavelets. Some toy examples illustrate the feasibility of these constructions and gives promising results. Now, the important points to be investigated are a way for cheaply computing the kernel and to infer the “best” wavelet basis from the data. These points are still open issues.

References


