Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 69 (2006) 2180–2192
NEUROCOMPUTING
www.elsevier.com/locate/neucom

Building meaningful representations for nonlinear modeling of 1d- and 2d-signals: applications to biomedical signals

R. Dubois[a], B. Quenet[a], Y. Faisandier[b], G. Dreyfus[a,*]

[a]*ESPCI-Paristech, Laboratoire d'Electronique, 10 rue Vauquelin, 75005 Paris, France*
[b]*ELA Medical C.A. La Boursidière, 92357 Le Plessis-Robinson Cedex, France*

Received 27 July 2004; received in revised form 25 July 2005; accepted 25 July 2005
Available online 23 February 2006
Communicated by J. Tin-Yau Kwok

Abstract

The paper addresses two problems that are frequently encountered when modeling data by linear combinations of nonlinear parameterized functions. The first problem is *feature selection*, when features are sought as functions that are *nonlinear in their parameters* (e.g. Gaussians with adjustable centers and widths, wavelets with adjustable translations and dilations, etc.). The second problem is the design of an *intelligible representation* for 1D- and 2D- signals with peaks and troughs that have a definite meaning for experts. To address the first problem, a generalization of the orthogonal forward regression method is described. To address the second problem, a new family of nonlinear parameterized functions, termed *Gaussian mesa functions*, is defined. It allows the modeling of signals such that each significant peak or trough is modeled by a single, identifiable function. The resulting representation is sparse in terms of adjustable parameters, thereby lending itself easily to automatic analysis and classification, yet it is readily intelligible for the expert. An application of the methodology to the automatic analysis of electrocardiographic (Holter) recordings is described. Applications to the analysis of neurophysiological signals and EEG signals (early detection of Alzheimer's disease) are outlined.

Keywords: Signal modeling; Nonlinear regression; Orthogonal forward regression; Feature selection; Holter; Electrocardiography; Electroencephalography

1. Introduction

Modeling a signal by a family of parameterized functions is particularly useful in a variety of fields such as pattern recognition, feature extraction, classification or modeling. It is a straightforward way of performing information compression: the finite set of parameters of the modeling function may be a sparse representation of the signal of interest.

Typical families of parameterized functions used for modeling are polynomials, wavelets, radial basis functions, neural networks, etc. For a given modeling problem, the choice between those families is based on such criteria as implementation complexity, sparsity, number of variables of the quantity to be modeled, domain knowledge. The latter factor is actually the driving force behind the methodology described in the present paper.

More specifically, the scope of this article is twofold: first, we address the problem of *feature selection*, i.e. the problem of finding the most appropriate set of functions within a given family of functions that are *nonlinear in their parameters*; the solution that we describe here is generic. The second purpose is more application-specific: the design of a *meaningful representation* for 1D or 2D- signals that exhibit bumps and/or troughs having specific meanings for the domain expert, i.e. the problem of finding a representation such that each bump or trough is modeled by a single, uniquely identifiable function. The intelligibility of the representation by the expert is especially important in the field of biological signal analysis: an application of our method to anomaly detection from electrocardiographic recordings is described (1D-signals), and an application to the

*Corresponding author. Tel.: +33 1 40 79 4541; fax: +33 1 47 07 1393.
E-mail address: Gerard.Dreyfus@espci.fr (G. Dreyfus).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2005.07.014

modeling of time-frequency maps of electrophysiology and electro-encephalography recordings (2D-signals) is outlined.

The first part of the paper is devoted to the description of generalized orthogonal forward regression (GOFR), an extension of the powerful orthogonal forward regression (OFR) method of modeling by parameterized functions that are linear with respect to their parameters. We show that OFR can be extended to modeling by functions that are nonlinear with respect to their parameters. We show that GOFR overcomes some important limitations of traditional OFR.

In the second part of the paper, we define *Gaussian mesa functions*, which are shown to be especially appropriate for modeling signals that exhibit positive and negative peaks, in such a way that each peak can be appropriately modeled by a single mesa function.

Finally, we describe an application of the methodology to the automatic analysis of long-term electrocardiographic recordings (Holter recordings). We first show how each positive or negative peak can be modeled by a single mesa function. Then we show how each function can be labeled, automatically and unambiguously, with the labels used routinely by experts, and how automatic discrimination between two types of heartbeats can be performed with that signal representation. As a final illustration, we outline an application of the methodology to time–frequency maps from electrophysiological and electroencephalographic recordings.

2. Orthogonal forward regression for feature selection

2.1. The feature selection problem

Let g_γ be a parameterized function and γ the vector of its parameters. Let $\Omega = \{g_\gamma\}_{\gamma\in\Gamma}$ be a family of such functions, where Γ is the set of the parameters. Note that the cardinality of Ω can be either finite or infinite.

Modeling a function f ($f \in L^2(\mathbb{R})$) with M functions, chosen from Ω, consists of finding a function $\tilde{f}$ that is a linear combination of M functions of Ω such that the discrepancy e_M between f and $\tilde{f}$ is as small as possible:

$$f = \sum_{\substack{i=1 \\ g_{\gamma_i}\in\Omega}}^{M} \alpha_i g_{\gamma_i} + e_M. \tag{1}$$

That problem amounts to estimating M parameter vectors $\{\gamma_i\}_{i=1..M}$ and M scalar parameters $\{\alpha_i\}_{i=1..M}$ to construct $\tilde{f}$. It can be solved in two steps:

- a *feature selection* step: in the set Ω, find the subset of M functions that are most relevant to the modeling of the signal of interest (see for instance [9,16]),
- an *optimization* step: find the parameters of the functions selected as relevant features at the previous step.

2.1.1. Optimization

In the optimization step, $\{\gamma_i, \alpha_i\}_{i=1..M}$ are estimated from training data, i.e. specific values $\{x_k\}_{k=1..N}$ of the variable (or vector of variables), for which measurements f_k of the signal were performed; the measurements are assumed to have additive zero-mean noise ε_k:

$f_k = f(x_k) + \varepsilon_k$. The set $\{(x_k, f_k)\}_{k=1..N}$ is called the training set.

The least squares cost function J is defined as:

$$J = \sum_{k=1}^{N} \left(f_k - \tilde{f}(x_k)\right)^2 = \sum_{k=1}^{N} \left(f_k - \sum_{\substack{i=1 \\ g_{\gamma_i}\in\Omega}}^{M} \alpha_i g_{\gamma_i}(x_k) \right)^2. \tag{2}$$

Eq. (2) can be also written in the following form, highlighting the modeling error $e_M(x_k)$ and the measurement noise ε_k:

$$J = \sum_{k=1}^{N} \left(\left(f_k - f(x_k)\right) + \left(f(x_k) - \tilde{f}(x_k)\right) \right)^2 = \sum_{k=1}^{N} (\varepsilon_k + e_M(x_k))^2. \tag{3}$$

The optimal model in the least squares sense $\tilde{f}$ is obtained by minimizing function J with respect to its parameters:

$$\tilde{f} = \sum_{\substack{i=1 \\ g_{\gamma_i}\in\Omega}}^{M} \alpha_i g_{\gamma_i} \tag{4}$$

$$\text{with } J\left(\{\alpha_i, \gamma_i\}_{i=1,..M}\right) = \min_{\alpha\in R, \gamma\in\Gamma}\left(J(\{\alpha, \gamma\})\right).$$

2.1.2. Feature selection

The minimization of J is a multivariable nonlinear optimization problem, which is usually solved by iterative algorithms such as the BFGS algorithm or the Levenberg–Marquardt algorithm (see for instance [12,15]). Being iterative, those algorithms require the choice of initial values of the parameters $\{\alpha_i, \gamma_i\}_{i=1,...M}$. Therefore, prior to the optimization step, the number M of functions must be chosen, together with the initial values of the M parameter vectors $\{\gamma_i\}$ and of the parameters $\{\alpha_i\}$.

For functions that are local in space, such as Gaussians, random initialization of the parameters (centers and variances) is not recommended, because many random initializations and optimizations may be required in order to find a satisfactory model. In such a case, a frequent strategy consists in choosing one Gaussian per observation of the training set, centered on that point in input space, and with arbitrary variance [14]. The main shortcoming of the above initialization is the fact that the number of selected functions (M) is not optimal: it is related to the number of examples, which may have no relation whatsoever to the complexity of the data to be modeled. The least-squares support vector machine (LS-SVM, also known as Ridge SVM) [5] starts with one function per example, and performs a selection depending on the complexity of the margin boundary, but the parameters of the RBF functions