Automatica 54 (2015) 340–347

Contents lists available at ScienceDirect

Automatica

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

Brief paper

An instrumental least squares support vector machine for nonlinear system identification☆

Vincent Laurain[a,b], Roland Tóth[c,1], Dario Piga[d], Wei Xing Zheng[e]

[a] Université de Lorraine, CRAN, UMR 7039, 2 rue Jean Lamour, 54519 Vandoeuvre-lès-Nancy Cedex, France

[b] CNRS, CRAN, UMR 7039, France

[c] Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands

[d] IMT Institute for Advanced Studies Lucca, Piazza San Ponziano 6, 55100 Lucca, Italy

[e] School of Computing and Mathematics, University of Western Sydney, Penrith NSW 2751, Australia

ARTICLE INFO

Article history:
Received 19 April 2013
Received in revised form
31 August 2014
Accepted 31 January 2015
Available online 2 March 2015

Keywords:
Support vector machines
Reproducing kernel Hilbert space
Instrumental variables
Nonlinear identification
Machine learning
Non-parametric estimation

ABSTRACT

Least-Squares Support Vector Machines (LS-SVMs), originating from Statistical Learning and Reproducing Kernel Hilbert Space (RKHS) theories, represent a promising approach to identify nonlinear systems via nonparametric estimation of the involved nonlinearities in a computationally and stochastically attractive way. However, application of LS-SVMs and other RKHS variants in the identification context is formulated as a regularized linear regression aiming at the minimization of the ℓ_2 loss of the prediction error. This formulation corresponds to the assumption of an auto-regressive noise structure, which is often found to be too restrictive in practical applications. In this paper, Instrumental Variable (IV) based estimation is integrated into the LS-SVM approach, providing, under minor conditions, consistent identification of nonlinear systems regarding the noise modeling error. It is shown how the cost function of the LS-SVM is modified to achieve an IV-based solution. Although, a practically well applicable choice of the instrumental variable is proposed for the derived approach, optimal choice of this instrument in terms of the estimates associated variance still remains to be an open problem. The effectiveness of the proposed IV based LS-SVM scheme is also demonstrated by a Monte Carlo study based simulation example.

1. Introduction

Support vector machines (SVMs) have been originally developed as a class of *supervised learning* methods in stochastic learning theory. Their original purpose was to provide efficient tools for data analysis and pattern recognition in classification problems and regression analysis (Schölkopf & Smola, 2002; Vapnik, 1998). SVMs have had a paramount impact on the *machine learning* field since their extension as a theoretical framework in that setting (Cristianini & Taylor, 2000). These methods also offer an attractive, so-called non-parametric way of data-driven dynamic modeling, i.e., *system identification*, especially in the nonlinear context. In that context, these approaches are part of the data-driven model learning avenue (Ljung, 2010; Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014; Pillonetto, Quang, & Chiuso, 2011), focusing on the paradigm of estimation of the targeted system without posing prior assumptions on its dynamical nature or the non-linearities involved. Most of the research interest regarding identification with SVMs has been dedicated to *nonlinear block models* so far, using various *least-square* SVM (LS-SVM) approaches where the original nonlinear estimation problem is posed as a linear regression (Falck, Pelckmans, Suykens, & De Moor, 2009; Goethals, Pelckmans, Suykens, & De Moor, 2005). In general, LS-SVMs are particular variations of the original support vector machine approach using a regularized ℓ_2 loss function instead of a so called ϵ-insensitive loss function on the prediction error of the model. A particular advantage of expressing both the regularization and the loss in the ℓ_2 norm is that the solution of the corresponding optimization problem is obtained by solving a system of

☆ This work was supported by the Netherlands Organization for Scientific Research (NWO, grant no.: 639.021.127) and by the French ministries of Foreign Affairs, Education and Research and the French-Dutch Academy (PHC Van Gogh project, n. 29342QL). The material in this paper was partially presented at the 50th IEEE Conference on Decision and Control, December 12–15, 2011, Orlando, FL, USA. This paper was recommended for publication in revised form by Associate Editor Alessandro Chiuso under the direction of Editor Torsten Söderström.

E-mail addresses: vincent.laurain@univ-lorraine.fr (V. Laurain), r.toth@tue.nl (R. Tóth), dario.piga@imtlucca.it (D. Piga), W.Zheng@uws.edu.au (W.X. Zheng).

[1] Tel.: +31 40 247 2655; fax: +31 40 243 4582.

http://dx.doi.org/10.1016/j.automatica.2015.02.017
0005-1098/

linear equations and an attractive trade-off between regularization bias and variance of the estimates is present (Goethals et al., 2005). LS-SVMs are also related to Kriging (Krige, 1966) in geostatistics and *Gaussian processes* (GPs) in machine learning, e.g., Frigola and Rasmussen (2013) and Kocijan, Girard, Banko, and Murray-Smith (2005), whose approaches can be seen as different variants of the *reproducing kernel Hilbert space* (RKHS) theory based function estimators. The relation between these methods is analyzed in Pillonetto et al. (2014) and Van Gestel et al. (2002).

A particular handicap of the variants of LS-SVMs (and also GPs) is that the used linear regression form under the ℓ_2 loss function corresponds to the assumption that all disturbances affecting the data-generating system can be expressed as a white noise disturbance on the equation error level, which can be seen as the assumption of a *nonlinear auto-regressive* noise structure. Such an assumption is often found to be too restrictive in practical applications. In the classical identification literature, significant research efforts have been devoted to achieve consistent estimation in case of rather general noise assumptions corresponding to the situations commonly encountered in practice (Ljung, 1999). To introduce the same generality of noise structures, some steps have been taken in the LS-SVM context such as the recurrent LS-SVM developed in Suykens and Vandewalle (2000) and the linear parametric noise model equipped SVM derived in Falck, Suykens, and De Moor (2010). However, the classical results in identification highlight that the chosen noise model, i.e., the assumed noise properties, plays an important role in the consistency of the estimates. Therefore, in the light of a non-parametric prior-free modeling objective, the question rises why we should bound ourself to *a priori* specified noise assumptions, especially in the general nonlinear context. For example, in the GPs related literature for LTI models, consistency under general noise conditions is established by identifying the one-step-ahead predictor of the output, which, due to linearity, allows factorization of high order linear regression based estimates to obtain estimates of the process and the noise dynamics without posing any priors on the noise (Pillonetto, Chiuso, & De Nicolao, 2011). However, in the nonlinear case, the loss of linearity of this predictor in the inputs and outputs prevents applicability of this methodology, allowing consistency only under restrictive assumptions, see Pillonetto et al. (2011). So, the important question that rises is how we can achieve similar generality in the nonlinear case.

By turning to the classical results, we can find that variants of linear regression based methods, e.g., *instrumental variable* (IV) approaches, have been developed to cope with realistic assumptions on the noise without specifying a direct parametrization or structure (Ljung, 1999). The strength of IV methods in the LTI case has been found in delivering consistent estimates independently of the chosen noise model assumption in a computationally attractive way (Young, 2011). Therefore, to extend consistency of non-parametric identification with LS-SVMs in the nonlinear case, in this paper, we consider the idea of introducing the IV scheme into the LS-SVM regression structure, which was first[2] proposed in Laurain, Zheng, and Tóth (2011). As a significant improvement of the initial scheme described in Laurain et al. (2011), in this paper, we provide a rigorous treatment of instrumental variables based LS-SVMs and showing the applicability of IV based techniques both in non-parametric identification and in regularized contexts. Furthermore, this contribution not only preserves the computationally attractive feature of the original approach by satisfying the *Mercer conditions*, but also provides unbiased estimates under general noise model structures/conditions; opening a large set of application areas for data-driven nonlinear model learning.

The paper is organized as follows: the considered problem setting and the motivation for improving the LS-SVM method are discussed in Section 2. In Section 3, the optimization problem associated with the IV-based, non-parametric model estimation is introduced together with its solution. This is followed by integrating the IV solution into the LS-SVM estimation scheme for nonlinear dynamic systems resulting in the IV-SVM method. In Section 4, the choice of the instrumental variables is discussed from the variance point of view together with the selection of kernel functions and tuning of the hyper parameters. To demonstrate the advantages of the IV-SVM, a Monte Carlo study in Section 5 is provided in which the identification of a nonlinear system under colored noise is analyzed. Finally, conclusions and some future directions of research are given in Section 6.

2. Problem description

To set the stage for the upcoming discussion, the considered identification problem is defined in this section.

2.1. The data-generating system

As an objective of the identification scenario, the data-driven estimation of a rather general class of nonlinear discrete-time systems is considered. For the sake of simplicity of the upcoming derivations, the system $\mathcal{S}_o$ is assumed to be *single-input single-output* (SISO). The behavior of $\mathcal{S}_o$ is described by the following difference equation

$$y(k) = f_o(x(k)) + v_o(k), \tag{1}$$

where $x(k) \in \mathbb{R}^{n_g}$ is a vector which, in the present identification context, is composed of the delayed values of the output and input signals of $\mathcal{S}_o$, y and u respectively:

$$x(k) = [y(k-1) \; \dots \; y(k-n_a) \; u(k) \; \dots \; u(k-n_b)]^\top,$$

with $n_g = n_a + n_b + 1$. $f_o : \mathbb{R}^{n_g} \to \mathbb{R}$ is assumed to be a bounded nonlinear function belonging to the set of real square integrable functions $\mathcal{L}_2(\mathbb{R}^\cdot, \mathbb{R})$. $v_o(k)$ is considered as a zero-mean, quasi-stationary stochastic noise process (not necessarily white), independent of u. Note that the general structure of the system defined by (1) can be used to describe usual block structures such as *Hammerstein* and/or *Wiener* systems by *a priori* restrictions of the structure of f_o, e.g.:

$$y(k) = \sum_{i=1}^{n_a} f_i^o\big(y(k-i)\big) + \sum_{j=0}^{n_b} g_j^o\big(u(k-j)\big) + v_o(k). \tag{2}$$

Formulation of (1) in the *multi-input multi-output* (MIMO) case is also available as shown in Goethals et al. (2005). Note that in case $v_o = e_o$, where e_o is white, (1) can be seen as a *nonlinear auto-regressive with exogenous input* (NARX) model (Sjöberg et al., 1995).

2.2. The modeling paradigm

To briefly discuss the concept behind the LS-SVM estimator and to develop the motivations for the proposed extension of this approach, let us consider the classical parametric estimation of (1), in which the nonlinearity is assumed to have an expansion (e.g., see Bai, 1998):

$$f_o\big(x(k)\big) = \sum_{i=1}^{n_H} \theta_i^o \phi_i\big(x(k)\big), \tag{3}$$

[2] Note that IV has also been applied to nonlinear systems in Lundgren and Sjöberg (2004). However, Lundgren and Sjöberg (2004) is not related to the current work as it only applies a parametric IV method to identify local LTI models of a nonlinear system around some operating conditions.