

## Chapter 3.19

# Soft–Constrained Linear Programming Support Vector Regression for Nonlinear Black– Box Systems Identification

**Zhao Lu**

*Tuskegee University, USA*

**Jing Sun**

*University of Michigan, USA*

### ABSTRACT

As an innovative sparse kernel modeling method, support vector regression (SVR) has been regarded as the state-of-the-art technique for regression and approximation. In the support vector regression, Vapnik (2000) developed the  $\varepsilon$ -insensitive loss function as a trade-off between the robust loss function of Huber and one that enables sparsity within the support vectors. The use of support vector kernel expansion provides us a potential avenue to represent nonlinear dynamical systems and underpin advanced analysis. However, in the standard quadratic programming support vector regression (QP-SVR), its implementation is more computationally expensive and enough model

sparsity cannot be guaranteed. In an attempt to surmount these drawbacks, this chapter focuses on the application of soft-constrained linear programming support vector regression (LP-SVR) in nonlinear black-box systems identification, and the simulation results demonstrates that the LP-SVR is superior to QP-SVR in model sparsity and computational efficiency.

### INTRODUCTION

Models of dynamical systems are of great importance in almost all fields of science and engineering and specifically in control, signal processing, and information science. A model is always only an approximation of a real phenomenon so that hav-

DOI: 10.4018/978-1-60960-195-9.ch319

ing an approximation theory which allows for the analysis of model quality is a substantial concern. A fundamental principle in system modeling is the Occam's razor arguing that the model should be no more complex than is required to capture the underlying systems dynamics. This concept, known as the parsimonious principle, which ensures the smallest possible model that explains the data, is particularly relevant in nonlinear model building because the size of a nonlinear model can easily become explosively large.

During the past decade, as an innovative sparse kernel modeling technique, support vector machine (SVM) has been gaining popularity in the field of machine learning and has been regarded as the state-of-the-art technique for regression and classification applications (Cristianini & Shawe-Taylor, 2000; Schölkopf & Smola, 2002; Vapnik, 2000). Essentially, SVM is a universal approach for solving the problems of multidimensional function estimation. Those approaches are all based on the Vapnik-Chervonenkis (VC) theory. Initially, it was designed to solve pattern recognition problem, where in order to find a decision rule with good generalization capability, a small subset of the training data, called the support vectors (SVs), are selected. Experiments showed that it is easy to recognize high-dimensional identities using a small basis constructed from the selected support vectors. Since the inception of this subject, the idea of support vector learning has also been applied to various fields, such as regression, density estimation, and linear operator equation, successfully. When SVM is employed to tackle the problems of function approximation and regression estimation, the approaches are often referred to as the support vector regression (SVR) (Smola & Schölkopf, 2004). The SVR type of function approximation is very effective, especially for the case of having a high-dimensional input space. Another important advantage for using SVR in function approximation is that the number of free parameters in the function approximation scheme is equal to the number of support vectors. Such

a number can be obtained by defining the width of a tolerance band, which can be implemented by using the  $\epsilon$ -insensitive loss function. Thus, the selection of the number of free parameters can be directly related to the approximation accuracy and does not have to depend on the dimensionality of the input space or other factors as that in the case of multilayer feedforward neural networks.

The  $\epsilon$ -insensitive loss function is attractive because unlike the quadratic and Huber cost functions, where all the data points will be support vectors, the SV solution can be sparse. In the realm of data modeling, the sparsity plays a crucial role in improving the generalization performance and computational efficiency. It has been shown that sparse data representations reduce the generalization error as long as the representation is not too sparse, which is consistent with the principle of parsimony (Ancona, Maglietta, & Stella, 2004; Chen, 2006).

For the purpose of modeling complex nonlinear dynamical systems by sparse representation, SVR was used in the context of nonlinear black-box system identification very recently (Chan, Chan, Cheung, & Harris, 2001; Drezet & Harrison, 1998; Gretton, Doucet, Herbrich, Rayner, & Schölkopf, 2001; Rojo-Alvarez, Martinez-Ramon, Prado-Cumplido, Artes-Rodriguez, & Figueiras-Vidal, 2006). Although it is believed that the formulation of SVM embodies the structural risk minimization principle, thus combining excellent generalization properties with a sparse model representation, data modeling practitioners have begun to realize that the capability for the standard quadratic programming SVR (QP-SVR) method to produce sparse models has perhaps been overstated. For example, it has been shown that the standard SVM technique is not always able to construct parsimonious models in system identification (Drezet & Harrison, 1998). A recent study (Lee & Billings, 2002) has compared the standard SVM and uniformly regularized orthogonal least squares (UROLS) algorithms using time series prediction problems, and has found that both methods have

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/soft-constrained-linear-programming-support/49424](http://www.igi-global.com/chapter/soft-constrained-linear-programming-support/49424)

## Related Content

---

### Introduction of Digital Watermarking

(2012). *Signal Processing, Perceptual Coding and Watermarking of Digital Audio: Advanced Technologies and Models* (pp. 14-19).

[www.irma-international.org/chapter/introduction-digital-watermarking/56057](http://www.irma-international.org/chapter/introduction-digital-watermarking/56057)

### Synthetic Video Generation for Evaluation of Sprite Generation

Yi Chen and Ramazan S. Aygün (2010). *International Journal of Multimedia Data Engineering and Management* (pp. 34-61).

[www.irma-international.org/article/synthetic-video-generation-evaluation-sprite/43747](http://www.irma-international.org/article/synthetic-video-generation-evaluation-sprite/43747)

### Navigating Through Video Stories Using Clustering Sets

Sheila M. Pinto-Cáceres, Jurandy Almeida, Vânia P. A. Neris, M. Cecília C. Baranauskas, Neucimar J. Leite and Ricardo da S. Torres (2011). *International Journal of Multimedia Data Engineering and Management* (pp. 1-20).

[www.irma-international.org/article/navigating-through-video-stories-using/58048](http://www.irma-international.org/article/navigating-through-video-stories-using/58048)

### I-Gate: Interperception - Get all the Environments

Rummenigge Dantas, Luiz Marcos Gonçalves, Claudio Schneider, Aquiles Burlamaqui, Ricardo Dias, Hugo Sena and Julio cesar Melo (2011). *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts* (pp. 397-413).

[www.irma-international.org/chapter/gate-interperception-get-all-environments/50601](http://www.irma-international.org/chapter/gate-interperception-get-all-environments/50601)

### K-Means Based Prediction of Transcoded JPEG File Size and Structural Similarity

Steven Pigeon and Stéphane Coulombe (2012). *International Journal of Multimedia Data Engineering and Management* (pp. 41-57).

[www.irma-international.org/article/means-based-prediction-transcoded-jpeg/69520](http://www.irma-international.org/article/means-based-prediction-transcoded-jpeg/69520)