Long-term prediction of nonlinear time series with recurrent least squares support vector machines

Indir Jaganjac

ArcelorMittal Zenica
Department of Iron Making
Bulevar Kralja Tvrtka 17/I
Bosnia and Herzegovina

Abstract. This paper is about applying recurrent least squares support vector machines (LS-SVM) on three ESTSP08 competition datasets. Least squares support vector machines are used as nonlinear models in order to avoid local minima problems. Then prediction task is re-formulated as function approximation task. Recurrent LS-SVM uses nonlinear autoregressive exogenous (NARX) model to build nonlinear regressor, by estimating in each iteration the next output value, given the past output and input measurements.

1 Introduction

Support Vector Machines (SVM) is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation [4]. It has been originally introduced within the context of statistical learning theory and structural risk minimization. These methods use quadratic programming to solve convex optimization problems [1]. Least Square Support Vector Machines (LS-SVM) are re-formulation to the standard SVM [6,7]. In this paper, recurrent least squares support vector machines are used as nonlinear models in order to avoid local minima problems [4,8]. The cost function is a regularized least squares function with equality constraints, leading to linear Karush-Kuhn-Tucker systems [4,7,8]. LS-SVMs are closely related to regularization networks and Gaussian processes [6,9]. Accurate prediction of nonlinear time series is very important in many fields: wind power systems, seismology, econometrics, industrial process automation systems, biomedicine, life sciences and etc. The main difficulty is lack of sufficient and necessary information for an accurate prediction. The challenge in the field of time series prediction is the long-term prediction, where typically more than 100 steps ahead ought to be predicted. Long-term prediction methods must solve many problems because of accumulation of errors, noise and perturbations from the environment. This paper is organized as follows. In Section 2, main principle of LS-SVM for nonlinear function estimation is presented. Section 3 presents how recurrent LS-SVM and nonlinear autoregressive exogenous (NARX) models are used for nonlinear regression and prediction. Section 4 presents ESTSP08 3 datasets and results. Section 5 concludes with some final remarks and pointers to further works.

2 Least-squares support vector machines for nonlinear function estimation

For a given training set of N data points $\{x_k, y_k\}$ with x_k as n-dimensional input and y_t as 1-dimensional output, feature space SVM models take the form [3,5]:

$$y(x) = \omega^T \varphi(x) + b,$$

where the nonlinear mapping φ(.) maps the input data into a higher dimensional feature space. In least-squares support vector machines (LS-SVM) for nonlinear function estimation, the following optimization problem is formulated:

$$\min_{\omega, e} J(\omega, e) = \tfrac{1}{2}\, \omega^T \omega + \gamma\, \tfrac{1}{2} \sum_{k=1}^{N} e_k^2,$$

subject to equality constraints:

$$y(x) = \omega^T \varphi(x) + b + e_k, \; k=1, \ldots, N$$

and the solution is:

$$h(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b$$

In the above equations, i refers to the index of a sample and $K(x, x_i)$ is the Kernel function defined as the dot product between the $\varphi(x)^T$ and $\varphi(x)$. In this paper, Gaussian kernels are used:

$$K(x, x_i) = \exp \{ ||x-x_i||^2 / \sigma^2 \}$$

The model hyperparameters σ and γ are trained and optimized according to [2,7,8].

3 Recurrent least-squares support vector machines

To predict more than 100 steps ahead values of time series, recurrent least-squares support vector machines can be used [3,5]. It uses the predicted values as known data to predict the next ones. The recurrent LS-SVM model can be constructed by first making one-step ahead prediction:

$$y_{t+1}' = f_1(y_t, y_{t-1}, \ldots, y_{t-M+1})$$

where M denotes the number of inputs and y_{t+1}' denotes predicted value. The regressor of the model is defined as the vector of inputs: $y_t, y_{t-1}, \ldots, y_{t-M+1}$. To predict the next value, the same model is used:

$$y_{t+2}' = f_1(y_{t+1}', y_t, y_{t-1}, \ldots, y_{t-M+1})$$

In this equation, the predicted value of y_{t+1}' is used instead of the true value, which is unknown. Then, for the H-steps ahead prediction, y_{t+2} to y_{t+H}' are predicted iteratively. When the regressor length M is larger than H, there are M-H real data in regressor to predict H_{th} step. When H exceeds M, all inputs are predicted values.

Nonlinear autoregressive exogenous (NARX) models are built based on nonlinear regression by estimating in each iteration the next output value, given the past output and input measurements. A dataset is converted into a new input (past measurements) by function *windowize*. Prediction is done by the function *predict*, iteratively in recurrent mode, and next output is computed, based on the previous predictions and starting values [2].

4 ESTSP08 times series datasets and results

4.1 Time Series 1

Function estimation is done for the 3_{rd} variable only, which has 354 data points. Then prediction is computed in recurrent mode for the next 18 values. The recurrent LS-SVM model is trained and fine-tuning of hyperparameters σ^2 and γ was done with cross-validation, according to [2].

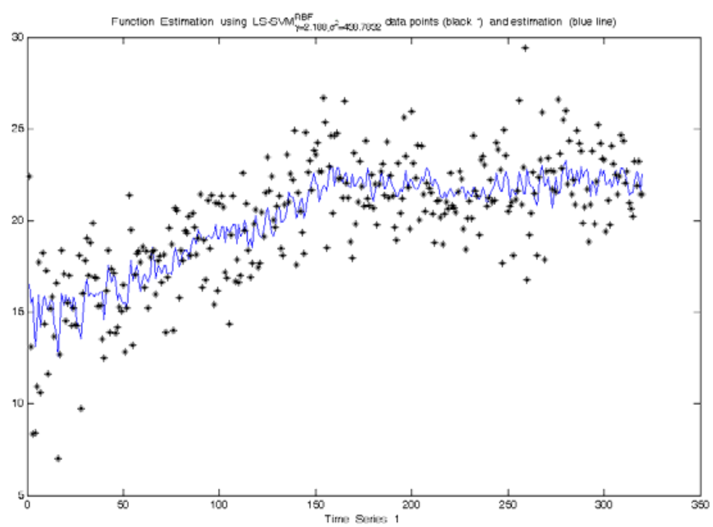

Figure 1: Function Estimation of Time Series 1

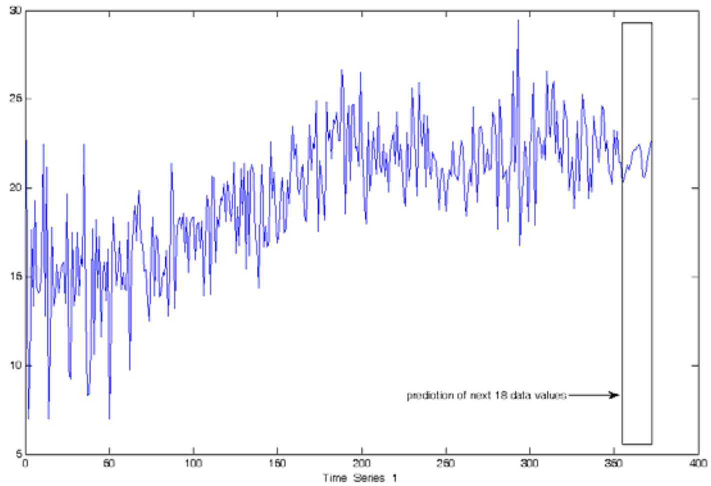

Figure 2: Prediction of next 18 data values for Time Series 1

4.2 **Time Series 2**

Nonlinear function estimation is done for the 1300 data points. Then prediction is computed in recurrent mode for the next 100 data values. The recurrent LS-SVM model is trained and fine-tuning of hyperparameters σ^2 and γ was done with cross-validation, according to [2].

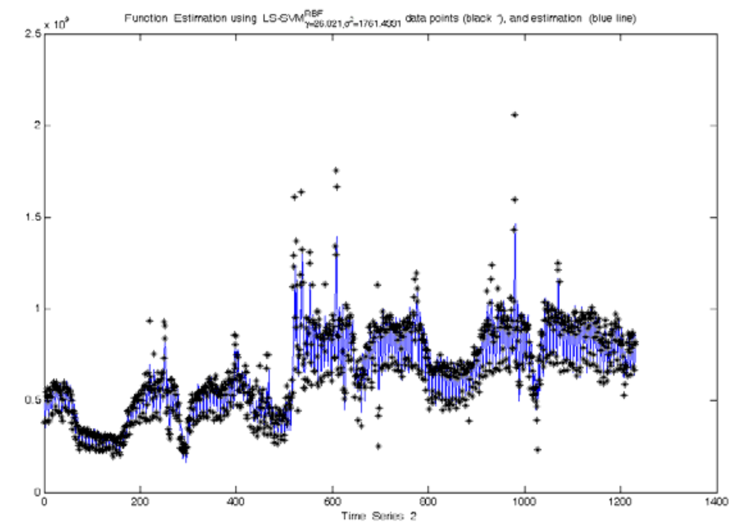

Figure 3: Function Estimation of Time Series 2

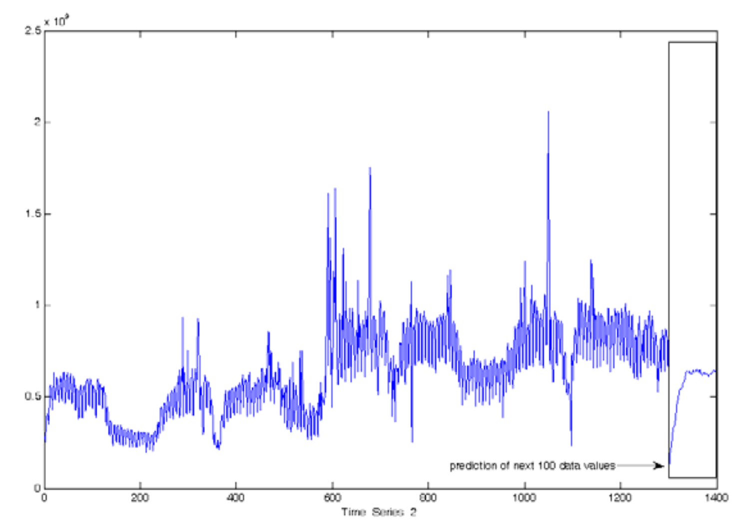

Figure 4: Prediction of next 100 data values for Time Series 2

4.3. **Time Series 3**

Nonlinear function estimation is done for the 3900 data points, which is sufficient for the accurate estimation. Then prediction is computed for the next 200 data values. Figure 5 shows nonlinear function estimation for 3900 data points and Figure 6 shows prediction of next 200 data values. The rationale for truncating the original dataset is that after testing various lower bounds, the best candidate lower bound converged to 3900 data points [10]. This lower bound captured all necessary fluctuations and dynamics in the original dataset. The recurrent LS-SVM model is trained and fine-tuning of hyperparameters σ^2 and γ was done with cross-validation, using Matlab and LS-SVMlab Toolbox. Details can be found in [2].

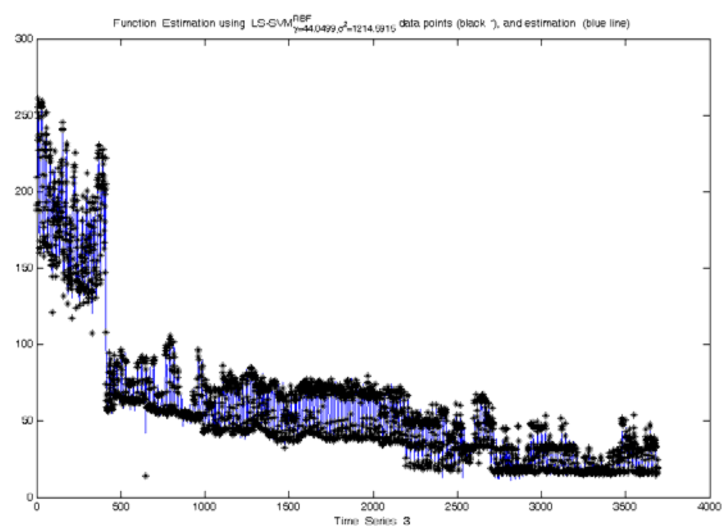

Figure 5: Function Estimation of Time Series 3

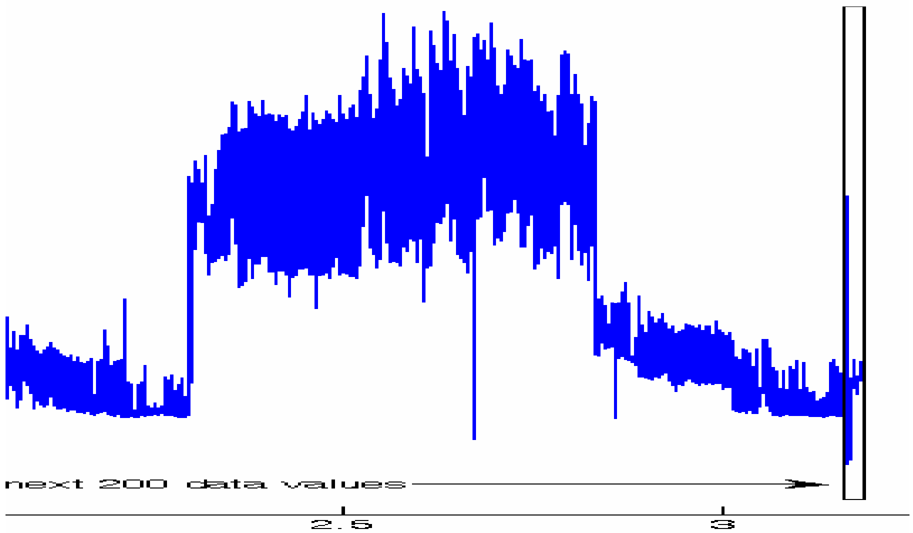

Figure 6: Prediction of next 200 data values for Time Series 3

5 Conclusion

This paper presents solutions for the long-term predictions of ESTSP08 datasets. LS-SVM was chosen for modeling nonlinear time series, because of its ability to avoid local minima problems. The prediction task was re-formulated as the nonlinear function estimation task. The predictions were computed using recurrent least squares support vector machines.

The recurrent least squares support vector machine iteratively predicts next output, based on the previous predictions and starting values. The time series 3 dataset was truncated to 3900 points, which was obtained as satisfactory lower bound for capturing all underlying fluctuations and dynamics in the original dataset.

In further works:

- research on parallelization of recurrent LS-SVM models will be done,
- on-line implementation of recurrent LS-SVM for continuous data analysis of industrial processes and measurements.

6 References

[1] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[2] K. Pelckmans, J.A.K. Suykens, T. Van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B. De Moor, J. Vandewalle, LS-SVM Toolbox User's Guide, ESAT-SCD-SISTA Technical Report 02-145, Katholieke Universiteit Leuven, February 2003.

[3] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, A. Lendasse, Methodology for long-term prediction of time series, Neurocomputing, 2007.

[4] J.A.K. Suykens, Support Vector Machines: A Nonlinear Modeling and Control Perspective, April, 2001.

[5] A. Lendasse, V. Wertz, G. Simon, M. Verleysen, Fast Bootstrap applied to LS-SVM for Long Term Prediction of Time Series, IJCNN '2004 proceedings, July 2004, IEEE pp. 705-710.

[6] C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[7] J.A.K. Suykens and J. Vandewalle, Recurrent Least Squares Support Vector Machines, IEEE Transactions on Circuits and Systems – I, Vol.47, No.7, July, 2000.

[8] J.A.K. Suykens, J.D. Brabanter, L. Lukas, J. Vandewalle, Weighted least squares support vector machines, robustness and sparse approximation, Neurocomputing 48, 2002.

[9] T. Poggio and F. Girosi, Networks for Approximation and Learning, Proceedings of the IEEE, Vol.78, No.9, September 1990.

[10] P. Denbigh, System Analysis & Signal Processing, Addison – Wesley, 1998.