A Novel Recursive Solution to LS-SVR for Robust Identification of Dynamical Systems

José Daniel de Alencar Santos[1]
Guilherme de Alencar Barreto[2]

[1]Federal Institute of Education, Science and Technology of Ceará, Department of Industry (IFCE)
[2]Federal University of Ceará, Department of Teleinformatics Engineering (UFC - DETI)

16th International Conference on Intelligent Data Engineering and Automated Learning - IDEAL'2015

October 2015

Agenda

Introduction

Motivation

- Least Squares Support Vector Regression (LS-SVR) is a widely used tool for time series forecasting, control and system identification.
- The Ordinary Least Squares (OLS) algorithm is too sensitive to the presence of outliers.
- Most of real-world data contains non-Gaussian noise.
- Lack of work on LS-SVR robust variants for system identification tasks.

Introduction

Objectives

1 - Propose a novel recursive and robust LS-SVR solution for regression problems.

2 - Evaluate the behavior of the proposed approach in system identification tasks, using infinite steps ahead prediction on both artificial and real data.

3 - Compare the performance of our method to some robust variants for LS-SVR models:

   a - WLS-SVR: Weighted Least Squares Support Vector Regression;

   b - IRLS-SVR: Iteratively Reweighted Least Squares Support Vector Regression.

Nonlinear System Identification

- Nonlinear autoregressive with exogenous inputs (NARX) model:

$$
\begin{aligned}
y_i &= m_i + \epsilon_i, \qquad m_i = g(\boldsymbol{x}_i), \qquad \epsilon_i \sim \mathcal{N}(\epsilon_i|0, \sigma_n^2), && (1) \\
\boldsymbol{x}_i &= [y_{i-1}, y_{i-2}, \cdots, y_{i-L_y}, u_{i-1}, u_{i-2}, \cdots, u_{i-L_u}]^T, && (2)
\end{aligned}
$$

  where $\boldsymbol{x}_i \in \mathbb{R}^P$ is the input vector, $u_i \in \mathbb{R}$ is the control input and $y_i \in \mathbb{R}$ is the output.
- After N instants, we have the dataset $\mathcal{D} = (\boldsymbol{x}_i, y_i)|_{i=1}^N = (\boldsymbol{X}, \boldsymbol{y})$, where $\boldsymbol{X} \in \mathbb{R}^{N \times P}$ is the regressor matrix and $\boldsymbol{y} = [y_1, \ldots, y_N]$.
- Given a new instant j, the prediction for test data follows

$$
\begin{aligned}
\hat{y}_j &= f(\boldsymbol{x}_j) + \epsilon_j, && (3) \\
\boldsymbol{x}_j &= [\hat{y}_{j-1}, \hat{y}_{j-2}, \cdots, \hat{y}_{j-L_y}, u_{j-1}, u_{j-2}, \cdots, u_{j-L_u}]^T, && (4)
\end{aligned}
$$

  where $\hat{y}_j$ is the j-th estimated noisy output.

Evaluated Models: LS-SVR

- For the estimation dataset $\{(\boldsymbol{x}_n, y_n)\}|_{n=1}^{N}$, the goal in a regression problem is to search for a function $f(\cdot)$ that takes the form

$$f(\boldsymbol{x}) = \langle \boldsymbol{w}, \varphi(\boldsymbol{x}) \rangle + b, \tag{5}$$

where $\varphi(\cdot) : \mathbb{R}^p \rightarrow \mathbb{R}^{p_h}$ is a nonlinear map, $\langle \cdot, \cdot \rangle$ denotes dot product, $\boldsymbol{w} \in \mathbb{R}^{p_h}$ is a vector of weights and $b \in \mathbb{R}$ is a bias.

- The parameter estimation problem leads to the minimization of

$$J(\boldsymbol{w}, e) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\frac{1}{2}\sum_{n=1}^{N} e_n^2, \tag{6}$$

subject to

$$y_n = \langle \boldsymbol{w}, \varphi(\boldsymbol{x}_n) \rangle + b + e_n, \quad n = 1, 2, \ldots, N \tag{7}$$

where $e_n = y_n - f(\boldsymbol{x}_n)$ is the n-th error and $C > 0$ is a regularization parameter.

Evaluated Models: LS-SVR

- After the application of Lagrangian function and the respective conditions for optimality, the dual problem corresponds to

$$\underbrace{\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{\Omega} + C^{-1}\mathbf{I} \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} b \\ \boldsymbol{\alpha}_0 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 0 \\ \boldsymbol{y}_0 \end{bmatrix}}_{\boldsymbol{y}}, \tag{8}$$

  where $\boldsymbol{\alpha}_0 \in \mathbb{R}^N$ is the vector of Lagrange multipliers, $\mathbf{\Omega} \in \mathbb{R}^{N \times N}$ is the kernel matrix ($\Omega_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$) and $k(\cdot, \cdot)$ is the kernel function.
- The solution for $\boldsymbol{\alpha}$ is obtained by OLS algorithm as

$$\boldsymbol{\alpha} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{y}. \tag{9}$$

- The resulting LS-SVR model for nonlinear regression is given by

$$f(\boldsymbol{x}) = \sum_{n=1}^{N} \alpha_n k(\boldsymbol{x}, \boldsymbol{x}_n) + b. \tag{10}$$

Evaluated Models: WLS-SVR

- The WLS-SVR model is obtained by the minimization of the functional

$$J(\boldsymbol{w}, e) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\frac{1}{2}\sum_{n=1}^{N} v_n e_n^2, \tag{11}$$

where $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ is a vector of weights associated with the error variables.

- The WLS-SVR solution is provided by solving the linear system $\boldsymbol{A}_v\boldsymbol{\alpha} = \boldsymbol{y}$, also using OLS algorithm

$$\boldsymbol{\alpha} = (\mathbf{A}_v^T\mathbf{A}_v)^{-1}\mathbf{A}_v^T\boldsymbol{y}, \tag{12}$$

where

$$\mathbf{A}_v = \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \boldsymbol{\Omega} + C^{-1}\mathbf{V} \end{bmatrix}, \quad \mathbf{V} = \text{diag}\left\{\frac{1}{v_1}, \ldots, \frac{1}{v_N}\right\}. \tag{13}$$

Evaluated Models: WLS-SVR

- The weights v_n are determined from Hampel weight function

$$v_n = \begin{cases} 1 & \text{if } |e_n/\hat{s}| \leq c_1, \\ \frac{c_2 - |e_n/\hat{s}|}{c_2 - c_1} & \text{if } c_1 < |e_n/\hat{s}| \leq c_2, \\ 10^{-4} & \text{otherwise,} \end{cases} \tag{14}$$

where $\hat{s} = \text{IQR}/1.349$ [1] is a robust estimate of the standard deviation of the error variables e_n, $c_1 = 2.5$ and $c_2 = 3.0$.

[1] IQR stands for InterQuantile Range, which is the difference between the 75th percentile and the 25th percentile.

Evaluated Models: IRLS-SVR

- The weighting procedure for WLS-SVR model can be repeated iteratively giving rise to IRLS-SVR model.
- At each iteration i, it is necessary to solve a linear system $\mathbf{A}_v^{(i)}\boldsymbol{\alpha}^{(i)} = \boldsymbol{y}$, where

$$\mathbf{A}_v^{(i)} = \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \boldsymbol{\Omega} + C^{-1}\boldsymbol{V}_i \end{bmatrix}, \quad \boldsymbol{V}^{(i)} = \text{diag}\left\{\frac{1}{v_1^{(i)}}, \ldots, \frac{1}{v_N^{(i)}}\right\}. \quad (15)$$

- The resulting model at the i-th iteration is given by

$$f^{(i)}(\boldsymbol{x}) = \sum_{n=1}^{N} \alpha_n^{(i)} k(\boldsymbol{x}, \boldsymbol{x}_n) + b^{(i)}. \quad (16)$$

- The stopping criterion is $\max_n(|\alpha_n^{(i)} - \alpha_n^{(i-1)}|) \leq 10^{-3}$.

The Proposed Approach: RLM-SVR

- The proposed method works with the same optimization problem as the standard LS-SVR model. Then, we have to solve the linear system

$$\underbrace{\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & \mathbf{\Omega} + C^{-1}\mathbf{I} \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} b \\ \boldsymbol{\alpha}_0 \end{bmatrix}}_{\boldsymbol{\alpha}} = \underbrace{\begin{bmatrix} 0 \\ \boldsymbol{y}_0 \end{bmatrix}}_{\boldsymbol{y}}. \tag{17}$$

- The idea is to solve (17) for the parameter vector $\boldsymbol{\alpha}$ by using the Recursive Least M-Estimate (RLM) [2] algorithm.
- For each vector $\boldsymbol{a}_n \in \mathbb{R}^{N+1}$ from the above matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n, \ldots, \boldsymbol{a}_{N+1}]$, we should apply the following recursive procedure

[2] The RLM rule is a robust variant of the standard Recursive Least Squares (RLS) algorithm, which uses M-estimators to handle outliers.

The Proposed Approach: RLM-SVR

1 - Calculate the a priori error: $e_n = y_n - \boldsymbol{\alpha}_{n-1}^T \boldsymbol{a}_n$, where $\boldsymbol{\alpha}_0 = \mathbf{0}$;

2 - Calculate the estimative of the error variance σ_n^2 as

$$\hat{\sigma}_n^2 = \lambda_e \hat{\sigma}_{n-1}^2 + c(1 - \lambda_e)\mathrm{med}(F_n), \tag{18}$$

where $0 \ll \lambda_e \leq 1$ is a forgetting factor, $\mathrm{med}(\cdot)$ is the median operator, $F_n = \{e_n^2, e_{n-1}^2, \dots, e_{n-N_w+1}^2\}$ and $c = 1.483(1 + 5/(N_w - 1))$;

3 - For the threshold parameters $\xi_1 = 1.96\hat{\sigma}_i$, $\xi_2 = 2.24\hat{\sigma}_i$ and $\xi_3 = 2.576\hat{\sigma}_i$, determine the Hampel's weight function as

$$q(e_n) = \begin{cases} 1 & 0 \leq |e_n| < \xi_1, \\ \frac{\xi_1 \mathsf{sign}(e_n)}{e_n} & \xi_1 \leq |e_n| < \xi_2, \\ \frac{\xi_1(\xi_3 - |e_n|)}{\xi_3 - \xi_2} \frac{\mathsf{sign}(e_n)}{e_n} & \xi_2 \leq |e_n| < \xi_3, \\ 0 & \text{otherwise}, \end{cases} \tag{19}$$

The Proposed Approach: RLM-SVR

4 - Apply the RLM algorithm using the following equations

$$\boldsymbol{S}_n = \lambda^{-1}(\boldsymbol{I} - \boldsymbol{g}_n \boldsymbol{a}_n^T)\boldsymbol{S}_{n-1}, \tag{20}$$

$$\boldsymbol{g}_n = \frac{q(e_n)\mathbf{S}_{n-1}\boldsymbol{a}_n}{\lambda + q(e_n)\boldsymbol{a}_n^T\mathbf{S}_{n-1}\boldsymbol{a}_n}, \tag{21}$$

$$\boldsymbol{\alpha}_n = \boldsymbol{\alpha}_{n-1} + (y_n - \boldsymbol{a}_n^T\boldsymbol{\alpha}_{n-1})\boldsymbol{g}_n, \tag{22}$$

where $\boldsymbol{S}_n$ is the inverse of the M-estimate correlation matrix of $\boldsymbol{a}_n$, $\boldsymbol{g}_n$ is the M-estimate gain vector and $0 \ll \lambda \leq 1$ is a forgetting factor;

5 - The cycle from the step 1 to 4, over all the data samples, should be repeated for $N_e > 1$ times. In this paper, we assume $N_e = 20$.

Simulations and Discussion

Datasets

- The artificial datasets were incrementally corrupted with a number of outliers equal to $2.5\%, 5\%$ e 10% of the estimation samples.
- A uniformly distributed value $U(-M_y, +M_y)$ was added to each randomly chosen sample, where M_y is the maximum absolute output value.

Dataset	estimation samples	test samples	noise
Artificial 1	150	150	$\mathcal{N}(0, 0.0025)$
Artificial 2	300	100	$\mathcal{N}(0, 0.29)$
Real dataset	512	512	?

Simulations and Discussion

Artificial 1

$$y_i = y_{i-1} - 0.5\tanh(y_{i-1} + u_{i-1}^3)$$

$u_i \sim \mathcal{N}(u_i|0,1), \quad -1 \leq u_i \leq 1,$ for both estimation and test data.

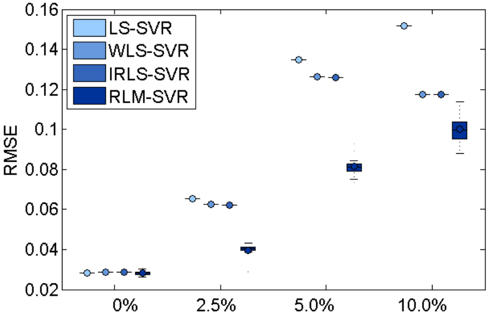

Figure: RMSE with test samples of Artificial 1 dataset in free simulation.

Simulations and Discussion

Artificial 2

$$y_i = \frac{y_{i-1} y_{i-2}(y_{i-1} + 2.5)}{1 + y_{i-1}^2 + y_{i-2}^2} + u_{i-1},$$

$u_i = U(-2, 2)$ for estimation data and $u_i = \sin(2\pi i/25)$ for test data.

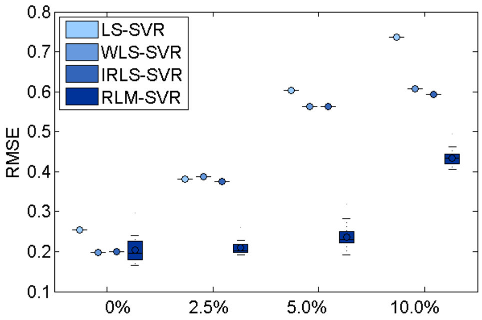

Figure: RMSE with test samples of Artificial 2 dataset in free simulation.

Simulations and Discussion

Wing flutter

- This data set is available in http://homes.esat.kuleuven.be/∼smc/daisy/daisydata.html.
- It corresponds to a SISO (Single Input Single Output) system, which input is highly colored.

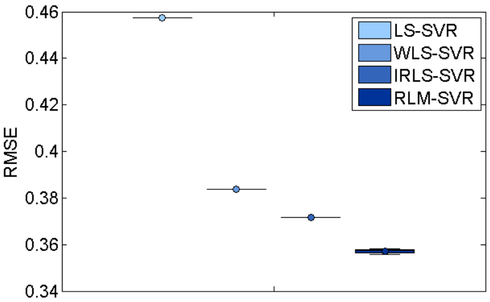

Figure: RMSE with test samples of wing flutter dataset in free simulation.

Conclusion

- It was proposed an outlier-robust recursive strategy to solve the parameter estimation problem of the standard LS-SVR model.
- The new approach, called RLM-SVR, is a robust variant of the RLS algorithm which uses M-estimators.
- Application in system identification tasks using infinite step ahead prediction in the presence of outliers.
- Performance comparison with two robust LS-SVR variants, WLS-SVR and IRLS-SVR.
- For all used datasets, the worst results achieved by the RLM-SVR model (maximum RMSE values) were better than the other methods.
- For outlier-free scenarios, the RLM-SVR model performed similarly to the others.

References

1 - Ljung, L.: System Identification Theory for the User. 2nd edn. (1999).

2 - Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: Least Squares Support Vector Machines. World Scientific Publishing, 1st edn. (2002).

3 - Suykens, J. A. K., De Brabanter, J., Lukas, L., Vandewalle, J.: Weighted Least Squares Support Vector Machines: Robustness and Sparse Approximation. Neurocomputing 48(1), 85-105 (2002).

4 - De Brabanter, K., Pelckmans, K., De Brabanter, J., Debruyne, M., Suykens, J. A., Hubert, M., De Moor, B.: Robustness of Kernel Based Regression: A Comparison of Iterative Weighting Schemes. In: Artificial Neural Networks - ICANN'2009. Springer 100–110, (2009).

5 - Zou, Y., Chan, S., Ng, T.: A Recursive Least M-Estimate (RLM) Adaptive Filter for Robust Filtering in Impulsive Noise. IEEE Signal Proc. Let. **7**(11), 324–326, (2000).