Weighted LS-SVM for Function Estimation Applied to Artifact Removal in Bio-signal Processing.

Alexander Caicedo and Sabine Van Huffel

***Abstract*—Weighted LS-SVM is normally used for function estimation from highly corrupted data in order to decrease the impact of outliers. However, this method is limited in size and big time series should be segmented in smaller groups. Therefore, border discontinuities represent a problem in the final estimated function. Several methods such as committee networks or multilayer networks of LS-SVMs are used to address this problem, but these methods require extra training and hence the computational cost is increased. In this paper a technique that includes an extra weight vector in the formulation of the cost function for the LS-SVM problem is proposed as an alternative solution. The method is then applied to the removal of some artifacts in biomedical signals.**

I. Introduction

BIOMEDICAL signals are normally corrupted by different artifacts and noise sources. These sources are related to physiological phenomena, electronic instrumentation or environmental conditions. The removal of the undesired disturbances in the signals is of utmost importance considering their impact in the processing algorithm used in later stages [1].

Due to the big variety of artifacts and their different impact in the signals, there exist different types of algorithms designed to reduce their influence under some specific condition. On the other hand, the noise is normally reduced by the use of filters of which the characteristics depends on the problem specifications.

Function estimation from noisy samples is closely related to signal denoising. In this framework the general problem is to find a function that minimizes the error ε between the predicted values *f(x)* and measured values *y*. Statistics (linear and nonlinear regression), neural networks and support vector machines among other methodologies have been used to solve this problem.

A Support Vector Machine (SVM) formulation is performed within the context of complex optimization theory, which involves a high computational cost [2]. Least Squares Support SVM (LS-SVM) is a reformulation of the SVM problem where inequality constraints are replaced by equality constraints and a quadratic loss function is taken for the error variables [3]. As a result, the problem simplifies to find the solution of a set of linear equations instead of solving a quadratic programming problem. However, the solution of the LS-SVM lacks 2 important properties: sparseness and robustness. In order to improve the robustness of the method a weighted version of the original LS-SVM algorithm is used [4].

Moreover, even if the complexity of formulation is reduced, large scale problems that involve a large number of samples can become memory and time consuming. By selecting a training subset from the original data or by the use of several LS-SVMs that later on are combined by means of a committee or neural network, this problem can be solved [3]. However, the model complexity is increased. Hence, joint LS-SVM models will be a better approach; but, big discontinuities are presented in the joint borders. In this study a new method that involves a modification of the weighted LS-SVM is exposed. With this method large scale problems can be solved as joint weighted LS-SVM models with minimal distortion in the joint borders.

This paper is organized as follows. Section II contains a brief introduction to LS-SVM for function estimation. Section III describes the weighted LS-SVM and the modification included to deal with border distortion. Section IV contains some experimental results used in the artifact removal in biomedical signals, while in section V conclusions are presented.

II. LS-SVM FOR FUNCTION ESTIMATION

Consider a training dataset $\{x_i, y_i\}_{i=1}^N$ where $x \in \Re^p$ is a p-dimensional input vector and $y \in \Re$ is the measured variable. Now consider the following model:

$$y(x) = \omega^T \varphi(x) + b \qquad (1)$$

ω is the unknown vector parameter, b is a bias term and $\varphi(x): \Re^p \rightarrow \Re^{p_h}$ represents a mapping (linear or nonlinear) to a high dimensional feature space. Then, the following optimization problem can be formulated:

$$\min_{\omega,b,e} J_P(\omega, e) = \frac{1}{2}\omega^T\omega + \gamma\frac{1}{2}\sum_{i=1}^{N} e_i^2$$

$$\text{Subject to } y_i = \omega^T\varphi(x_i) + b + e_i \qquad (2)$$

e represents the estimation error and γ is a regularization constant that represents a trade-off between smoother solution and residuals.

A. Caicedo is with the Department of Electronic Engineering ESAT-SCD, Katholieke Universiteit Leuven, Belgium (phone: +32 16 321067 ; fax: +32 16 321970; e-mail: alexander.caicedodorado@esat.kuleuven.be).

S. Van Huffel is with the Department of Electronic Engineering ESAT-SCD, Katholieke Universiteit Leuven, Belgium (e-mail: sabine.vanhuffel@esat.kuleuven.be).

Construct the Lagrangian:

$$\ell(\omega,b,e,\alpha) = J_P(\omega,e) - \sum_{i=1}^{N}\alpha_i\{\omega^T\varphi(x_i)+b+e_i-y_i\} \quad (3)$$

α_i are the Lagrange multipliers. Given the following conditions for optimality:

$$\begin{aligned}
&\frac{\partial\ell}{\partial\omega}=0 \quad\rightarrow\quad \omega=\sum_{i=1}^{N}\alpha_i\varphi(x_i)\\
&\frac{\partial\ell}{\partial b}=0 \quad\rightarrow\quad \sum_{i=1}^{N}\alpha_i=0\\
&\frac{\partial\ell}{\partial e_i}=0 \quad\rightarrow\quad \alpha_i=\gamma e_i\\
&\frac{\partial\ell}{\partial\alpha_i}=0 \quad\rightarrow\quad \omega^T\varphi(x_i)+b+e_i-y_i=0
\end{aligned} \quad (4)$$

and eliminating ω and e, the following solution is obtained:

$$\begin{bmatrix}0 & 1_v^T\\ 1_v & \Omega+I/\gamma\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix} \quad (5)$$

$1_v=[1...1]$ and $\Omega_{il}=\varphi(x_i)\varphi(x_l)=K(x_i,x_l)$ is the kernel matrix obtained by applying the kernel trick.

By replacing ω from (4) into (1), the resulting LS-SVM model becomes:

$$y(x)=\sum_{i=1}^{N}\alpha_i K(x,x_i)+b \quad (6)$$

A detailed discussion about the advantages and disadvantages of the LS-SVM can be found in [3].

III. WEIGHTED LS-SVM

In case of outliers and noise with non-Gaussian distribution the performance of the LS-SVM is affected. In order to reduce the impact of these outliers a weighted version of the LS-SVM algorithm is used.

The weighted LS-SVM algorithm first computes an unweighted LS-SVM and calculates the errors e_i. Then the standard deviation of the errors $\hat{s}$ is computed in order to identify the outliers that affect the model performance. Based on $\hat{s}$ a weight vector v is defined. Reference [5] presents the following formulation for the weight vector v:

$$v_i=\begin{cases}1 & if \quad |e_i/\hat{s}|\le c_1\\ \dfrac{c_2-|e_i/\hat{s}|}{c_2-c_1} & if \quad c_1\le|e_i/\hat{s}|\le c_2\\ 10^{-4} & otherwise\end{cases} \quad (7)$$

c_1 and c_2 are constant with typical values of 2.5 and 3 respectively. A new LS-SVM problem where the errors are weighted is formulated. The formulation of the problem is given by:

$$\min_{\omega,b,e} J_P(\omega,e)=\frac{1}{2}\omega^T\omega+\gamma\frac{1}{2}\sum_{i=1}^{N}v_i e_i^2$$
$$\text{Subject to } y_i=\omega^T\varphi(x_i)+b+e_i \quad (8)$$

Applying the Lagrangian and conditions for optimality the problem simplifies to:

$$\begin{bmatrix}0 & 1_v^T\\ 1_v & \Omega+V_\gamma\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix} \quad (9)$$

V_γ is a diagonal matrix given by:

$$V_\gamma=diag\left(\frac{1}{\gamma v_1},...,\frac{1}{\gamma v_N}\right) \quad (10)$$

Due to the lack of sparseness in the LS-SVM the number of support vectors is equal to the number of data points, moreover, (9) represents a system of (N+1)x(N+1) equations, with N equals to the number of data points. The higher this number the heavier the computational load. To restrict the latter, long signal recordings cannot be processed. Yet, in large scale problems, normally the data is segmented in M (consecutive or no) parts each of which is used to train a LS-SVM. Then all the resulting models are combined by means of a committee or a neural network. The results are satisfactory but the computational cost is increased. Although it would be better to join all estimated functions from M consecutive segments, due to the lack of shared information between the LS-SVM models the joint output will present high discontinuities at the borders. In order to address this problem, another formulation of the weight vector is used.

Consider the following problem:

$$\min_{\omega,b,e} J_P(\omega,e)=\frac{1}{2}\omega^T\omega+\gamma\frac{1}{2}\sum_{i=1}^{N}\frac{v_i}{\mu_i}e_i^2$$
$$\text{Subject to } y_i=\omega^T\varphi(x_i)+b+e_i \quad (11)$$

μ represents a new weight vector of the form presented in

Fig. 1. After applying the Lagrangian and the conditions for optimality the problem simplifies to:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \mathrm{M}_\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \qquad (12)$$

M_γ is a diagonal matrix with diagonal components given by:

$$\mathrm{M}_\gamma = diag\left(\frac{\mu_1}{\gamma_1}, ..., \frac{\mu_N}{\gamma_N} \right) \qquad (13)$$

This definition of the new weight vector guarantees that the data points located at the borders will have priority in the error minimization. It is important to note that the lowest value given to this function should be greater than 0 in order to avoid numerical problems in (11) and (12).

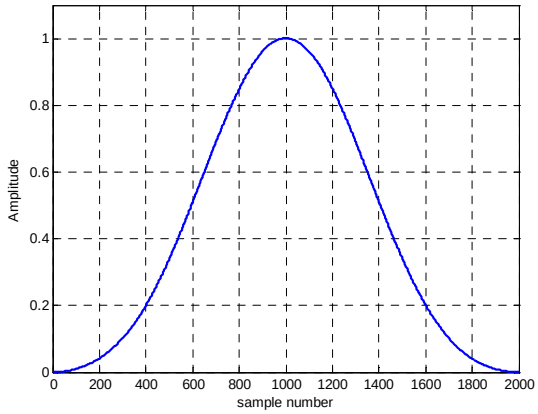


Fig. 1. A blackman window was used as the extra weight vector μ included in the LS-SVM formulation. The minimum value of the window by definition is 0; however, an offset of 1e-3 was used to avoid numerical problems in the problem formulation (11). This window guarantees that priority will be given to the reduction of the error in the borders; hence, discontinuities are reduced when the output of the submodels is merged.

Whereas in this paper the vector μ proposed reduces the discontinuities at the borders for the joint LS-SVM sub-models in the large scale function estimation, the same approach can be taken using other shapes for the weight vector μ in order to strengthen the fit in another region or regions of interest.

IV. EXPERIMENT

In this section the results of applying the weighted LS-SVM algorithm to remove artifacts in real life biomedical signals are presented.

A. Data

Long-term recordings of 6-72 hours of premature infants with need for intensive care were used. The signals measured were: Arterial Oxygen Saturation (SaO_2) measured continuously by pulse oximetry, Mean Arterial Blood Pressure (MABP), measured by an indwelling arterial catheter and Near Infrared Spectroscopy (NIRS) signals such as Haemoglobin differences (HbD) (measured by the Critikon Cerebral Oxygenation Monitor 2001), Regional Cerebral Oxygen Saturation (rSO_2) (INVOS4100, Somanetics Corp.), and Tissue Oxygenation Index (TOI) (NIRO300, Hamamatsu). The signals were measured simultaneously in the first days of life and were sampled with a sampling period of 3s. These signals are affected by huge artefacts due to movement, displacement of electrodes but mostly big changes in baseline due to blood samples extraction or disconnected electrodes. These drops can be interpreted as outliers in the time series and their elimination is the main goal for applying the proposed algorithm.

B. Method

In order to obtain the LS-SVM model for the function estimation the following steps were performed:

1) Kernel selection: the model presented in (1) requires a mapping of the input data to a higher dimensional space, this mapping is given by the kernel function selected $K(x_i, x_l)$. The kernel selection defines the hyper-parameters that together with γ need to be tuned. In this example, the Radial Basis Function (RBF) kernel is used. The RBF kernel is expressed as follows:

$$K(x, x_i) = e^{-\left(\frac{|x - x_i|_2^2}{\sigma^2} \right)} \qquad (14)$$

σ is the kernel bandwidth.

2) Training and validation dataset: In order to tune the model a training and a validation dataset are needed. A segment of length N=2220 free of artifacts from the original signal is selected, due to computational problems the value of N should be lower than 3000 samples (in a computer running windows with intel® Core™2 Quad CPU Q9400 @ 2.66GHz processor). The data is then divided in 10 segments, each segment contains N/10 samples. The criterion used to extract the samples from the original data is the same as the one presented for the fixed size LS-SVM algorithm in [2].

3) Model selection: In order to select the best hyper-parameters 10-fold cross-validation was used. 9 of the 10 segments are used for training and the 10^{th} segment is used for validation, the procedure is repeated until each segment has been used for validation, the validation error is then calculated as the mean value of the 10 validation errors calculated. The hyper-parameters selected are the ones who

minimize the validation error.

C. Test results

In Fig. 2, results for function estimation for a segment of 500 samples contaminated with artifact using the normal joint weighted LS-SVM model (8) and the newly proposed joint weighted LS-SVM model (11) are presented. It can be seen that the performance of both methods is similar in the segments free of artifacts; however, the newly proposed method follows the signal dynamics closer in the interval shown. This effect is due to the extra weight function used, as the time interval presented in Fig. 2 corresponds to the output of the model where the signal is weighted with the first 500 samples of μ in Fig. 1; due to the shape of this vector, data points that are located in this range will have a lower error than the normal method. In intervals of time where the weight vector μ reaches its maximum value 1 both methods yield the same performance. This shows that the selection of an appropriated weight vector μ can reduce the effects due to regions highly contaminated by noise in the final estimated function.

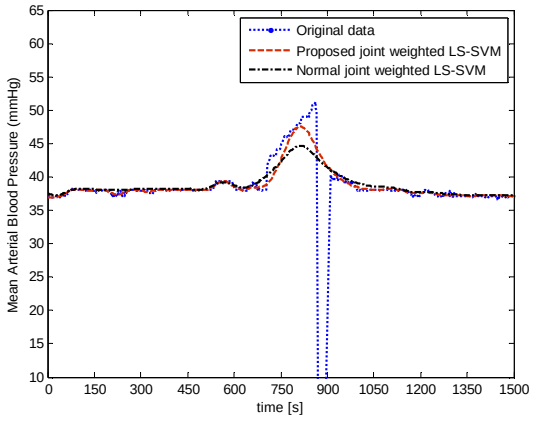


Fig. 2. Illustration of the performance of the normal joint weighted LS-SVM and the newly proposed joint weighted LS-SVM for MABP signal. The hyper-parameter values selected after training were $\gamma = 100.5$, and $\sigma = 438$.

In Fig. 3 the distortion in the joint points is shown. As can be seen, in the normal method a high distortion is presented at the borders, this discontinuity is caused by joining two estimated functions from two different weighted LS-SVMs. In the proposed method this discontinuity is reduced by forcing the estimated function to minimize the error at the borders. However, if an outlier is present in those points the effect of the weight vector μ will be reduced by the weight vector v as is expressed in (11). In order to avoid this situation it is preferred adjust the segment so as to move the discontinuities away from the borders and then perform the fitting. In practice this can be done by selecting thresholds or by applying a robust LS-SVM and localizing the outliers in the error function as explained in section III.

V. Conclusion

In this paper an extra weight vector for the weighted LS-SVM algorithm is proposed. This new vector is able to strengthen the fit in a selected region or regions of interest such as the borders. The method was applied to denoise and remove artifacts in recordings of MABP, SaO2 and NIRS signals, and was able to reduce the discontinuities presented in the estimated function due to joining the submodels outputs.

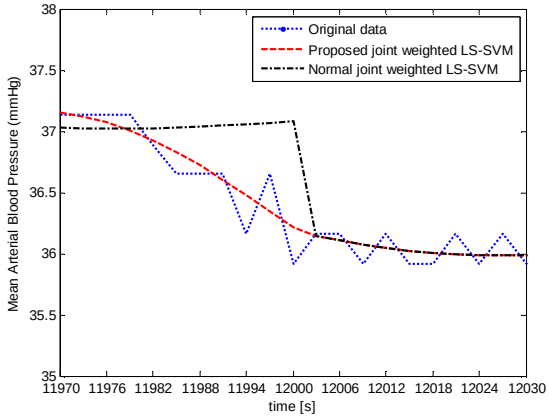


Fig. 3. Illustration of the joint models performance. As can be seen the normal joint weighted LS-SVM produces high discontinuities, while the newly proposed joint weighted LS-SVM reduces their influence.

Acknowledgment

Research supported by the Research Council KUL: GOA-AMBioRICS, GOA-MANET, CoE EF/05/006 Optimization in Engineering (OPTEC), IDO 05/010 EEG-fMRI, IDO 08/013 Autism, IOF-KP06/11 FunCopt, several PhD/postdoc & fellow grants; by FWO projects G.0519.06 (Noninvasive brain oxygenation) G.0302.07 (SVM), G.0341.07 (Data fusion), G.0427.10N (Integrated EEG-fMRI) research communities (ICCoS, ANMMM); and the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, `Dynamical systems, control and optimization', 2007-2011); ESA PRODEX No 90348 (sleep homeostasis) EU: FAST (FP6-MC-RTN-035801), Neuromath (COST-BM0601).

References

[1] R. M. Rangayyan. "*Biomedical Signal Analysis A Case-Study Approach*", J. Wiley &Sons, New york, 2002.

[2] N. Cristianini, and J. Shawe-Taylor. "*An introduction to Support Vector Machines*", Cambridge University Press.

[3] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle "*Least Squares Support Vector Machines*", World Scientific, 2005.

[4] J.A.K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "*Weighted least squares support vector machines: robustness and sparse approximation*", *Neurocomputing*, 2002. pp. 85-105.

[5] P. J. Rousseeuw, A. Leroy "*Robust Regression and Outlier Detection*", John Wiley & Sons, New York, 2003.