SYSTEM IDENTIFICATION BASED ON MULTI-KERNEL LEAST SQUARES SUPPORT VECTOR MACHINES (MULTI-KERNEL LS-SVM)

Mounira Tarhouni, Kaouther Laabidi, Moufida Lahmari-Ksouri
Unit of Research Analysis and control of systems (ACS, ENIT), BP 37, Le Belvedere 1002, Tunis, Tunisia

Salah Zidi
LAGIS (USTL, Lille), Villeneuve d’Ascq, 59650 Lille, France

Keywords: Nonlinear system identification, Least Squares Support Vector Machines (LS-SVM), Multi-kernel function, Multi model, Weighted function.

Abstract: This paper develops a new approach to identify nonlinear systems. A Multi-Kernel Least Squares Support Vector Machine (Multi-Kernel LS-SVM) is proposed. The basic LS-SVM idea is to map linear inseparable input data into a high dimensional linear separable feature space via a nonlinear mapping technique (kernel function) and to carry out linear classification or regression in feature space. The choice of kernel function is an important task which is related to the system nonlinearity degrees. The suggested approach combines several kernels in order to take advantage of their performances. Two examples are given to illustrate the effectiveness of the proposed method.

1 INTRODUCTION

In the field of automatic control, obtaining an accurate model of dynamical system is an important task that can influence the overall performance of the control loop (Lauer, 2008). System identification aims at finding such an appropriate model of the system based on preliminary equations and input/output data. The model can be classified on three types: White Box models, Grey Box models and Black Box models, the last one is considered when no physical insight is available nor used (Lauer, 2008).

Recently, a new kind of learning machine called Support vector machines (SVM) has been presented and used for classification (Zidi et al., 2006), (Diosan et al., 2007), (ElFerchichi et al., 2009), (Tarhouni et al., 2010) and for regression (Lauer, 2008). The basic idea is to map linear inseparable input data into a high dimensional linear separable feature space via a nonlinear mapping technique (kernel function) and to carry out linear classification or regression in feature space.

Recent developments in the literature on the SVM have emphasized the need to consider multiple kernels based learning method (Lanckriet et al., 2004), (Zhejing et al., 2004), (Sonnenburg et al., 2006) and (Diosan et al., 2007). This idea reflects the fact that practical learning problems often involve multiple, heterogeneous data sources. (Diosan et al., 2007) have developed an Evolved Combined Kernels (ECKs). They have considered a combination of multiple kernels and they have used a genetic algorithm (GA) for evolving these weights. This proposed approach is used for solving classification problems. They have compared their results against those obtained by (Lanckriet et al., 2004) with a combined kernel learnt with convex methods (CCKs).

A variant of SVM called Least Squares Support Vector Machines (LS-SVM) method has been developed as an attractive tool for identification of complex non linear systems (Suykens and Vandewalle, 1999). This method is a widely known simplification of the SVM, allowing to obtain the model by solving a system of linear equations instead of a hard quadratic programming problem involved by the standard SVM (Valyon and Horvath, 2005). It is empirically found that the LS-SVM method yields a similar performance compared to the classical SVM on classification problems, and often outperforms SVM in the regression case.
2 LEAST SQUARES SUPPORT VECTOR MACHINES (LS-SVM)

(Suykens and J.Vandewalle, 1999) have presented the LS-SVM approach, in which the following function $\varphi$ is used to approximate the output system as follows:

$$y = < w, \varphi(x) > + b$$  \hspace{1cm} (1)

where $x \in \mathbb{R}^m$ are the input data, $y \in \mathbb{R}$ are the output data and $\varphi(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^p$ is the nonlinear function that maps the input space into a higher dimension feature space.

LS-SVM approach was defined in the following constrained optimization problem:

$$\begin{align*}
\min_{w,b,e} & J_p(w,e) = \frac{1}{2} ||w||^2 + \frac{C}{2} \sum_{i=1}^{N} e_i^2 \\
\text{subject to:} & \quad y_i = < w, \varphi(x_i) > + b + e_i, \quad k = 1,2,\ldots,N
\end{align*}$$  \hspace{1cm} (2)

Where $e_i$ are slack variable and $C$ is a regularization factor.

The corresponding Lagrangian involving the dual variables $\alpha_i$ is described by relation (3).

$$L(w,b,e,\alpha) = J_p(w,e) - \sum_{i=1}^{N} \alpha_i( < w, \varphi(x_i) > + b + e_i - y_i)$$  \hspace{1cm} (3)

Where $\alpha_i$ are lagrange multipliers. The optimality conditions are:

$$\begin{align*}
\frac{\partial L}{\partial w} &= 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\
\frac{\partial L}{\partial b} &= 0 \Rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\
\frac{\partial L}{\partial e_i} &= 0 \Rightarrow \alpha_i = C e_i, \quad k = 1,\ldots,N
\end{align*}$$  \hspace{1cm} (4)

By replacing the expression of $w$ and $e_i$ in the optimization problem constraints (2), We obtain:

$$y_i = \sum_{j=1}^{N} \alpha_j \varphi(x_j) \varphi(x_i) > + b - \frac{1}{C} \alpha_i$$  \hspace{1cm} (5)

The parameters are recovered as solution of the following linear system:

$$\begin{bmatrix} 0 \\ I \end{bmatrix}^T \Omega + C^{-1} \frac{\partial L}{\partial \varphi} = \begin{bmatrix} 0 \\ y \end{bmatrix}$$  \hspace{1cm} (6)

Where $\Omega = \varphi(x_i)^T \varphi(x_j) = k(x_i,x_j)$ is the kernel matrix.

Using LS-SVM approach the output model will be given by relation (7).

$$y(x) = \sum_{i=1}^{N} \alpha_i k(x,x_i) + b$$  \hspace{1cm} (7)

Where $\alpha_i, b \in \mathbb{R}$ are the solution of equation (6), $x_i$ is training data and $x$ is the new input vector.
The LS-SVM model output is based on lagrange multipliers and kernel function $K(x_i, x)$. The lagrange multipliers were generated by resolving the equation system (6). These lagrange multipliers depend also on the kernel matrix $K(x_i, x_j)$. Hence, the choice of this function is a crucial task. In this note, the proposed kernel is obtained by combining a linear kernel and a nonlinear one that will be used successively to identify the linear and the nonlinear parts as shown Figure 2.

$$y(k) = y(k-1) y(k-2) (y(k-1) + 4.5) \over 1 + y(k-1)^2 + y(k-2)^2 + u(k-1)$$

(11)

With an input excitation signal chosen as a multiple square signal represented in figure 3.

We generate a 1000 patterns, the first 500 samples are used as training test and the remaining 500 samples are used to validate the obtained model. The fitness criterion (Bemporad et al., 2005):

$$FIT = (1 - \|\hat{y} - y\| / \|y - \bar{y}\|) \times 100\%$$

(12)

is used to measure the similarity between the real output system $y = [y(1) \cdots y(N)]^T$ and the estimated output $\hat{y} = [\hat{y}(1) \cdots \hat{y}(N)]^T$. $\bar{y}$ is the average of measurements $y(k), k = 1, \cdots, N$. Where $N$ is the number of available measurement points.

The training LS-SVM algorithm has as input the vector $x = [u(k) u(k-1) y(k-1) y(k-2) y(k-3)]^T$ and as output $y(k)$. Figure 4 shows the system output.
Parameter C is taken equal to 5 and we obtain a FIT criterion equal to 82.40%. To compare the obtained results using linear kernel with these obtained using the Gaussian one. We have considered a gaussian kernel: \( K(x, y) = \exp(-\frac{1}{2\sigma^2} (x-y)^2) \)

Its kernel parameter is selected by cross validation. The bandwidth \( \sigma \) value is equal to 1 and the regularization parameter C is also, taken equal to 5. Then, the FIT converges to 80.95%. 

The obtained identification result are presented in the figure 7. We notice that the real system output and the model output are overcome. The FIT criterion is taken equal to 88.64% which is better than the criterion obtained with the last two methods.

### 3.1 Identification of Multimodel System based on Multi-kernel LS-SVM

When the complexity of the plant increases the traditional methods of modeling prove, often, incompetent to generate a global model likely to give an account of all the system characteristics. The multimodel approach was developed to bring an answer to these problems (Messaoud et al., 2009), (Zouari et al., 2008). This approach consists to represent the process by several models.

The multi model identification based on our proposed method is shown in the following diagram 8:

The idea proposed in this case is to construct a library of models. Every operating domain is identified using the proposed Multi-kernel LS-SVM approach. The parameters of the kernel function and the weighting factors are different from model to another. Once, all models were selected we pass to the validation test. In this level, an online computation of different validities degrees of each elaborated base’s models is necessary. The validity degrees evaluate the
pertinence of each local model to describe the global plant behaviour. In the literature this step is based on the residue calculation (Delmotte, 1997), (Messaoud et al., 2009). The multi-model output is obtained either by switching between the models of the library, or by using a technique of fusion according to the models validities. Thus, the multi-model output is a switching between all these elementary laws.

To illustrate the effectiveness of the suggested approach, we consider an example of a multi-model system. The process is composed by the association of two under different structure systems represented by following non linear discrete time input/output representations.

\[ S_1 : y(k) = 0.8y(k-1) - 0.12y(k-2)^2 + 1.2u(k-1) - 0.4u(k-2)^2 \]  

\[ S_2 : y(k) = 0.8y(k-1) - 0.1y(k-2)^2 + 1.4u(k-1) \]  

The first mode described by relation (13) is considered for 400 first iterations, the second mode given by relation (14) is considered for remaining iterations. Random input, generated with uniform distribution, is applied to the process. The input vector of training LS-SVM algorithm is \( x = [u(k)u(k-1)y(k-1), y(k-2), y(k-3)]^T \).

To evaluate the proposed results, we modify the chosen parameters \( (C, \sigma, \text{cst}) \) in order to select the optimal values. Finally, the given method based in the combination of the two kernels succeed to converge with a minimum identification error compared with the two last methods.

![Figure 9: Identification based on LS-SVM with linear kernel.](image)

![Figure 10: Identification based on LS-SVM with Gaussian kernel.](image)

![Figure 11: Multi-kernel LS-SVM identification.](image)

The results are mentioned in the table (1). This table corresponds to the results of the first sub-system \( (S_1) \) and their obtained identification results are shown in figures 9, 10 and 11. LS-SVM parameters of model with linear kernel are chosen as \( C = 2, \sigma = -10 \) and \( n = 1 \). And those of model with Gaussian kernel are \( \sigma = \frac{1}{\sqrt{2}}, C = 2 \) which correspond to the optimal parameters. The obtained considered "FIT" criterion converges to 91.29% after the combination of these two kernel functions.

Figure 11 shows that the system output and the model output are overcome and the identification error is acceptable. These obtained results demonstrate the effectiveness of our provided idea in identification problems. The same procedure is applied in the second subsystem and the proposed multi kernel method demonstrate once again, his performances.

The two obtained multi kernel LS-SVM model were constructed. The global system behaviour switches between the different models.

To evaluate identification performances obtained with multi-model based on multi-kernel LS-SVM approach, we have considered the following sequence \( "S_1;S_2;S_3" \). The obtained identification results are given by the figure 12.

The considered criterion "FIT" converges to 95.47%

<table>
<thead>
<tr>
<th>(C)</th>
<th>(2\sigma^2)</th>
<th>(\text{cst})</th>
<th>(\text{FIT} )% (\text{gaussien})</th>
<th>(\text{FIT} )% (\text{linear})</th>
<th>(\text{FIT} )% (\text{combined})</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>120</td>
<td>87.192</td>
<td>74.8552</td>
<td>88.8552</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>110</td>
<td>82.2450</td>
<td>75.0399</td>
<td>87.5464</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>88.5845</td>
<td>74.6180</td>
<td>91.0016</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-10</td>
<td>89.5432</td>
<td>77.4000</td>
<td>91.0016</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-100</td>
<td>88.6217</td>
<td>77.9769</td>
<td>90.3726</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>84.6485</td>
<td>76.4695</td>
<td>88.8563</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>-10</td>
<td>85.7563</td>
<td>75.0650</td>
<td>90.0809</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>88.5900</td>
<td>75.0690</td>
<td>89.9968</td>
</tr>
</tbody>
</table>
which is an acceptable performance.

4 CONCLUSIONS

A multi kernel LS-SVM identification method is proposed in this paper. The suggested idea consists of training an LS-SVM algorithm with multiple kernel functions. A linear kernel combined with the Gaussian one have been chosen. These two weighted combined kernels were demonstrate a good performances in the identification context compared with the traditional LS-SVM method which uses unique kernel function.

REFERENCES


