Microwave Characterization Using Least-Square Support Vector Machines

Tarik Hacib[1], Yann Le Bihan[2], Mohamed Rachid Mekideche[1], Hulusi Acikgoz[2], Olivier Meyer[2], and Lionel Pichon[2]

[1]Laboratoire d'études et de modélisation en électrotechnique, Faculté des Sciences de l'Ingénieur, Université de Jijel, 18000 Jijel, Algérie

[2]Laboratoire de Génie Electrique de Paris, SUPELEC, UMR 8507 CNRS, UPMC Université Paris 06, Université Paris-Sud 11, 91192 Gif-sur-Yvette Cedex, France

This paper presents the use of the least-square support vector machines (LS-SVM) technique, combined with the finite element method (FEM), to evaluate the microwave properties of dielectric materials. The LS-SVM is a statistical learning method that has good generalization capability and learning performance. The FEM is used to create the data set required to train the LS-SVM. The performance of LS-SVM model depends on a careful setting of its associated hyper-parameters. Different tuning techniques for optimizing the LS-SVM hyper-parameters are studied: cross validation (CV), genetic algorithms (GA), heuristic approach, and Bayesian regularization (BR). Results show that BR provides a good compromise between accuracy and computational cost.

Index Terms—**Bayesian inference, least-square support vector machines (LS-SVM), microwave characterization, optimization techniques.**

I. Introduction

THE problems of the determination of dielectric constant ε' and loss factor ε'' of dielectric materials are usually formulated and solved as optimization problems, so iterative methods are commonly used to solve these kinds of problems. These methods involve solving a well-behaved forward problem in a feedback loop. Numerical methods such as finite element method (FEM) can be used to represent the forward process. However, iterative methods using a numerically based forward model are computationally expensive. In this situation, neural networks (NN) and other machine learning tools such as support vector machines (SVM) are a good alternative to iterative methods [1].

SVM are a recent powerful machine learning method. They are developed on the basis of statistical learning theory. Compared to NN, SVM do not require to define, more or less arbitrary, a number of hidden neurons. They exhibit good generalization capabilities thanks to an intrinsic tradeoff between fitting of the training data and model complexity. Nevertheless, the SVM adjustment is obtained using quadratic programming methods, which can be time-consuming and difficult to implement.

Generally speaking, SVM have several hyper-parameters. If these ones are not correctly chosen, performances will not be satisfactory. Most of the papers rely on cross validation (CV) for the tuning of SVM hyper-parameters [2], [3]. Global optimization techniques such as genetic algorithm (GA) and simulated annealing are also used in a few cases [4].

In this paper, we study a new inversion method for the evaluation of the microwave properties of dielectric materials (complex permittivity) from the admittance measured at the discontinuity plane of a coaxial open-ended probe. The method is based on the FEM and a least-squares support vector machines (LS-SVM) scheme.

LS-SVM are a reformulation of the standard SVM that uses equality constraints (instead of the inequality constraints implemented in standard SVM) and a least-squares error term to obtain a linear set of equations in a dual space [5]. Thus, LS-SVM allow to reduce the learning cost.

Different methods CV, GA, heuristic approach, and Bayesian regularization (BR) are evaluated for the tuning of the LS-SVM parameters. They are compared in terms of speed, accuracy, and computational complexity.

The FEM provides the data set required for the training of LS-SVM. A data set consists of input (complex admittance, frequency) and output $(\varepsilon', \varepsilon'')$ pairs.

II. LS-SVM for Function Estimation

Given a training set $(x_k, y_k)_{k=1}^{N}$ with input data $x_k \in R^n$ and output data $y_k \in R$, the LS-SVM model for nonlinear function approximation is represented in feature space as

$$y(\boldsymbol{x}) = \boldsymbol{w}^T \varphi(\boldsymbol{x}) + b. \tag{1}$$

Here, the nonlinear function $\varphi(.) : R^n \rightarrow R^{n_k}$ maps the input space to a higher-dimension feature space. b is a bias term and $\boldsymbol{w} \in R^n$ is the weight vector. The optimization problem consists in minimizing

$$\boldsymbol{J}(\boldsymbol{w}, \boldsymbol{e}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \gamma\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{e}_i^2. \tag{2}$$

Subject to the equality constraints

$$y_i = \boldsymbol{w}^T\varphi(x_i) + b + \boldsymbol{e}_i \qquad i = 1, \ldots, N \tag{3}$$

where the fitting error is denoted by $\boldsymbol{e}_i$. The hyper-parameter γ controls the tradeoff between the smoothness of the function y and the accuracy of the fitting. The solution is obtained after constructing the Lagrangian

$$L(\boldsymbol{w}, b, \boldsymbol{e}, \alpha) = \boldsymbol{J}(\boldsymbol{w}, \boldsymbol{e}) - \sum_{i=1}^{N}\alpha_i\left\{\boldsymbol{w}^T\varphi(x_i) + b + \boldsymbol{e}_i - y_i\right\} \tag{4}$$

Manuscript received December 23, 2009; accepted February 07, 2010. Current version published July 21, 2010. Corresponding author: T. Hacib (e-mail: tarik.hacib@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMAG.2010.2043657

0018-9464/$26.00 © 2010 IEEE

where α_i are Lagrangian multipliers. Application of the conditions for optimality yields the following linear system:

$$\begin{bmatrix} 0 & 1^T \\ 1 & \boldsymbol{\Omega} + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{y} \end{bmatrix} \tag{5}$$

where $\boldsymbol{y} = [y_1, \ldots, y_N], 1 = [1, \ldots, 1], \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]$, and $\boldsymbol{\Omega}_{il} = K(x_i, x_l)$ is the kernel matrix.

The resulting LS-SVM model for nonlinear function estimation becomes

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

where α_i, b constitute the solution to the linear system, and $K(x_i, x_l) = \varphi(x_i)^T \varphi(x_l)$ is the so-called kernel function. The most usual kernel functions are linear, polynomial, and above all radial basis function (RBF) [5]. In this paper, RBFs are used

$$K(\boldsymbol{x}, x_i) = exp\left(-\frac{\|\boldsymbol{x} - x_i\|_2^2}{\sigma^2}\right) \tag{7}$$

where σ is a constant defining the kernel width.

There is no doubt that the efficient performance of the LS-SVM model involves an optimal selection of the hyper-parameters σ and γ [included in (2)]. This can be done using one of several tuning techniques presented in the next section.

III. LS-SVM Hyper-Parameters Optimization

In general, there are roughly two classes of methods for LS-SVM hyper-parameters estimation (tuning).

1) *Systematic search*: In practice, most researchers have so far used cross validation.
2) *Optimization algorithms*: Potentially, we can use global or local optimization techniques such as genetic or gradient algorithms.

Each of these techniques has potential advantages and drawbacks. In this section, different tuning techniques applied to LS-SVM modeling are presented. An approach based on the Bayesian inference is also introduced.

A. Cross Validation

One of the most popular techniques for evaluating a set of parameter values is the CV [6]. In CV, the training set T is divided up into M partitions $(T_1, T_2, \ldots, T_M)$. For each hyper-parameter setting, the LS-SVM model is trained M times, where during each time one of the M subsets is held out while the remaining $(M-1)$ subsets are used to train the model. Then, the held-out subset (validation set) is used to estimate the performance of the trained model. The mean squared error of these M trials is used to estimate the generalization capability of the model with the considered hyper-parameter value. The hyper-parameter value that yields the lowest generalization error (lower error on the validation set) is then chosen. When more than one parameter needs to be tuned, the combined settings of all of the parameters can be evaluated using CV in the same way. When M is equal to the number of training samples in T, the result is leave-one-out cross validation (LOO-CV). LOO-CV can be described as being desirable but computationally highly expensive.

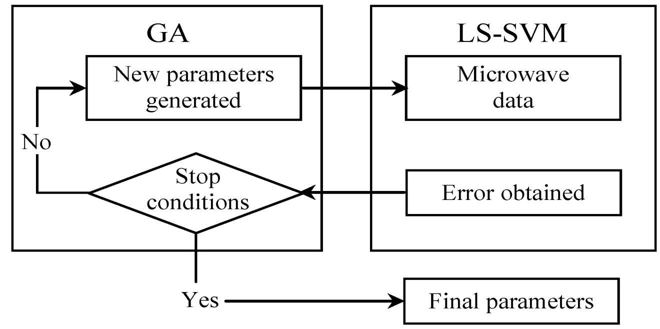

Fig. 1. GA technique with LS-SVM algorithm.

B. Genetic Algorithm

GA comprises a powerful stochastic optimization technique based on the processes of evolution theory. This method is reported to be suitable for a good approximate of global maximum or minimum value. A GA involves using three operators: reproduction, crossover, and mutation. In the context of LS-SVM, a GA produces sets of individuals, which represent the LS-SVM hyper-parameters γ and σ [4]. Each resulting LS-SVM is trained, and its performances evaluated by the mean of a root mean squared error $(RMSE)$ calculated on a validation set. Parents of the next generation are selected according to a fitness function that depends on the $RMSE$. Individuals with larger fitness value have greater possibility of being selected as parents. The fitness function is defined as

$$fitness = \frac{1}{RMSE(\sigma, \gamma)}. \tag{8}$$

Maximizing the fitness value corresponds to minimizing the prediction error. When a termination criterion is met, the individual having the best fitness defines the optimal hyper-parameters. Fig. 1 shows the implementation of GA technique in our microwave application.

C. Bayesian Regularization

BR was initially introduced in the field of neural networks. The method was then implemented in the context of LS-SVM [7]. It is based on the maximization of the posterior distribution of the LS-SVM parameters. For this purpose, we assume that the approximation error of the LS-SVM is Gaussian, white, and centered [8]. The prior distribution (i.e., without knowledge of the training set) of the LS-SVM internal parameters $\boldsymbol{w}$ and b is also assumed as Gaussian, white, and centered. Indeed, small-magnitude parameters will lead to a smooth fitting allowing a good generalization capability. The parameter inference is done in three distinct levels. First, the posterior distribution of the internal parameters $\boldsymbol{w}$ and b is maximized. This leads to minimize (2). The hyper-parameter γ is then deduced in a second level. The third level consists in an optimal model selection that leads in the context of LS-SVM with RBF kernel functions to determine the value of the width σ of the kernels. For the sake of brevity, the whole mathematical development of the Bayesian inference is not provided in this paper.

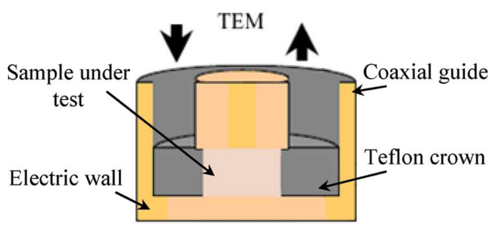

Fig. 2. SuperPol measuring cell.

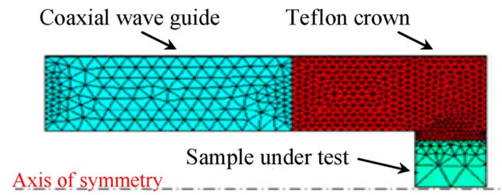

Fig. 3. View of the meshed measurement cell.

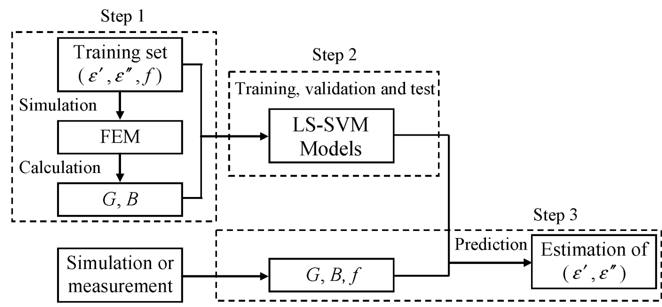

Fig. 4. Microwave properties parameter extracting procedure.

D. Hybrid Approach

A hybrid approach (HA) based on heuristic and GA was also implemented to tune the hyper-parameters γ and σ of LS-SVM models.

The implementation of the HA methodology is a two-step procedure. The first step consists in the determination of the Gaussian kernel width σ. Different heuristics have been proposed for this task. We used the following, which was already proposed in the framework of RBF NN [9]:

$$\sigma = \frac{d_{\max}}{\sqrt{2N}} \tag{9}$$

where $d_{\max}$ is the maximum distances between all input training data and N is the number of samples of this training data set. The subjacent idea is to obtain a partial overlapping between the different Gaussians corresponding to the samples of the training set in order to ensure interpolation capabilities. In the second part of the HA methodology, a GA seeks an optimal value of the hyper-parameter γ.

IV. Measurement Setup and Numerical Method

The characterization cell implemented in this work, called SuperPol, consists of a junction between a coaxial waveguide and a circular guide, which is filled in an inhomogeneous way. This means that the material under test is located in the continuity of the inner conductor and that it is held up by a Teflon crown (Fig. 2). The whole device is connected to an impedance analyzer.

This configuration may be used from low frequencies until only a few gigahertz using the coaxial waveguide GR900 ($\text{inner diameter} = 6.2$ mm, $\text{outer diameter} = 14.28$ mm).

The modeling of this configuration is done by using the FEM in order to generate the data sets required by the LS-SVM.

The problem is expressed in terms of the electric field $\boldsymbol{E}$, which satisfies the following harmonic wave equation:

$$\boldsymbol{curl}\left[-\frac{1}{i\omega\mu}\boldsymbol{curl}\ \boldsymbol{E}\right] - i\omega\varepsilon\boldsymbol{E} = 0 \tag{10}$$

where ω is the pulsation and $\varepsilon(\varepsilon = \varepsilon_0(\varepsilon' - i\varepsilon''))$ and μ are the permittivity and the permeability, respectively. ε_0 is the permittivity of the free space. Second-order 3D tetrahedral edge finite elements are used. Thanks to the symmetry of the system, only an angular sector of the geometry is meshed (Fig. 3).

V. Implementation and Evaluation of the LS-SVM

In this part, we summarize the steps of the LS-SVM implementation (Fig. 4):

1) generation of the samples of the data sets (training, validation, and test sets) and preprocessing of them (centring and normalization);
2) training, validation (except for BR, which does not involve validation), and test of the LS-SVM models (apply the different tuning algorithms presented in Section III to LS-SVM);
3) utilization of the designed LS-SVM for microwave data inversion.

Two single-output LS-SVM corresponding to the two estimated quantities ε' and ε'' are used.

Depending on the optimization algorithm used, there are some operational parameters that should be chosen properly for better performance and efficiency.

In CV, the number of subsets M in the training data set needs to be set. We implemented the *tenfold* CV $(M = 10)$, which is the most commonly used. γ is varying between 10 and 10^{10}, and σ between 10^{-2} and 5×10^2, where the two intervals are divided into 25 values logarithmically distributed. For GA, some parameters have to be determined prior to use LS-SVM model. For instance, population size, range of parameters, selection, and mutation operators have to be selected correctly. The range of variation of σ and γ is the same as for CV, a normalized representation is used to represent these parameters, uniform crossover and Gaussian mutation are used, and an initial population of size 30 is built randomly. With HA, only γ is determined by GA with the same range and with the same values than GA. BR has the attractive feature that it does not need any hyper-parameter predetermination.

VI. Results

LS-SVM obtained with the different optimization techniques are compared in terms of accuracy and time cost (time required for the elaboration of the LS-SVM model). ε' is varying between 1 and 100, and ε'' between 0 and 80, whereas the measurement frequency is from 1 MHz to 1.8 GHz. Figs. 5 and 6 show the time required for the model elaboration and the accuracy obtained on the test set (containing data not used in the elaboration of the

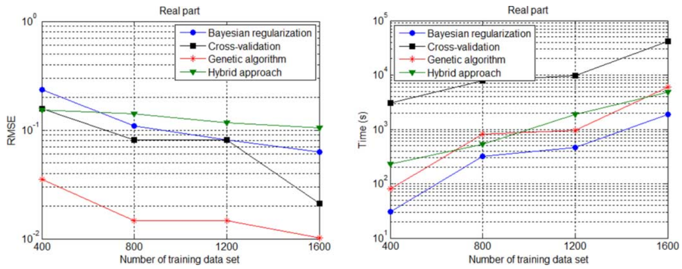

Fig. 5. Error and CPU time on the test set versus the number of training set data (evaluation of ε').

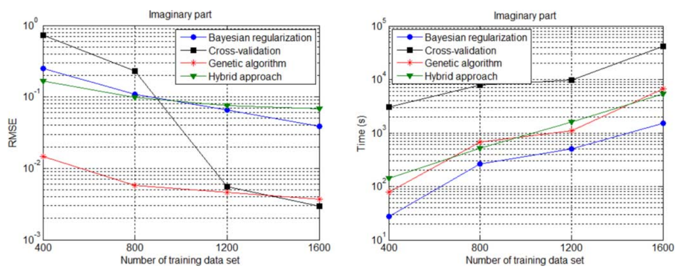

Fig. 6. Error and CPU time on the test set versus the number of training set data (evaluation of ε'').

LS-SVM, unlike the training and validation sets) for different sizes of training set.

At first sight, it can be observed that the LS-SVM with BR outperform the other methods in term of time cost, though leading to an acceptable accuracy. BR appears to be a powerful algorithm avoiding any predetermination of hyper-parameters. Both GA and heuristic-GA hybrid methodology are stochastic, flexible, and unlikely to get trapped in local minima. On the other hand, they are quite slow. Regarding our application, it was observed that they have almost the same time performance, though GA appeared to be more accurate. CV is the slowest approach. In this work, it appears also as being occasionally less accurate than GA. The choice of a finer step in the variation of the hyper-parameters should allow to improve the accuracy of CV at the expense of the time cost. Nevertheless, the simplicity of implementation of CV leads it to be a usual choice in most SVM problems.

As a measure of effectiveness, the optimized LS-SVM by BR was tested on ethanol data brought from experiments. LS-SVM results are compared to results obtained from a time-consuming iterative inversion method. This last method consists in using a direct model in an iterative procedure aiming to reduce the difference between the measured observation (complex admittance) and the calculated one. Measurements have been carried out by using an Agilent 4291A impedance analyzer on an ethanol sample whose dielectric characteristics are known. The thickness of the sample under test is 2.9 mm.

The following graph (Fig. 7) shows the good agreement between the LS-SVM results and those obtained by the iterative inversion procedure. Furthermore, the results fit with those found in the scientific literature.

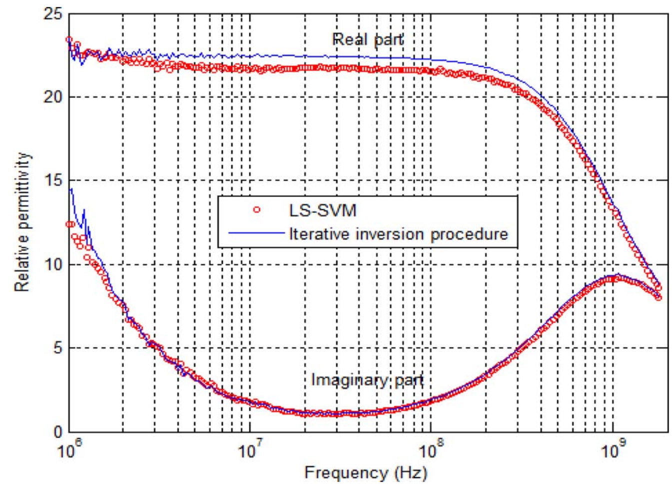

Fig. 7. Permittivity evolution of the ethanol obtained by LS-SVM and iterative inversion method.

VII. Conclusion

The SVM technique shows interesting attributes such as universal approximation capability, generalization capability, and sound theoretical foundations. In this paper, the applications of LS-SVM, an effectiveness kind of SVM, have been presented. However, the efficient performance of the LS-SVM model involves a correct choice of the kernel and regularization parameters, which can be done using different techniques. Various approaches for tuning the LS-SVM hyper-parameters were presented, and their performances were evaluated. As a result, it was observed that LS-SVM with Bayesian regularization, though more complex to program, show a good cost–accuracy compromise. It outperformed other tuning techniques including cross validation, genetic algorithms, and heuristic. Further works will concern the optimization of the structure of the training data set.

References

[1] S. Osowski, K. Siwek, and T. Markiewicz, "MLP and SVM networks-a comparative study," in *Proc. 6th NSPS*, Finland, 2004, pp. 37–40.

[2] L. P. Wang, *Support Vector Machines: Theory and Application*. Berlin, Germany: Springer, 2005.

[3] Y. Wu, Z. Tang, Y. Xu, Y. Guo, and B. Zhang, "Support vector regression for measuring electromagnetic parameters of magnetic thin-film materials," *IEEE Trans. Magn*, vol. 43, no. 12, pp. 4071–4075, Dec. 2007.

[4] C. Hsu, C. Wu, S. Chen, and K. Peng, "Dynamically optimizing parameters in support vector regression: An application of electricity load forecasting," in *Proc. HICSS*, 2006, vol. 2, p. 30c.

[5] J. A. K. Suykens, "Nonlinear modelling and support vector machines," in *Proc. IEEE IMTC*, Hungary, 2001, vol. 1, pp. 287–294.

[6] D. Wilson and T. Martinez, "Combining cross-validation and confidence to measure fitness," in *Proc. IJCNN*, 1999, vol. 2, pp. 1409–1414.

[7] J. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle, "Financial time series prediction using least squares support vector machines within the evidence framework," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 809–821, Jul. 2001.

[8] H. Acikgoz, Y. Le Bihan, O. Meyer, and L. Pichon, "Microwave characterization of dielectric materials using Bayesian neural networks," *Prog. Electromagn. Res. C*, vol. 3, pp. 169–182, 2008.

[9] L. Oukhellou, "Paramétrisation et Classification de Signaux en Contrôle non Destructif: Application à la Reconnaissance des Défauts de Rails par Courants de Foucault," Ph.D. dissertation, University Paris XI, Paris, France, 1997.