UNIVERSITY
OF TRENTO

DIPARTIMENTO DI INGEGNERIA E SCIENZA DELL'INFORMAZIONE

38123 Povo – Trento (Italy), Via Sommarive 14
http://www.disi.unitn.it

ON THE USE OF SVM FOR ELECTROMAGNETIC SUBSURFACE SENSING

A. Boni, M. Conci, A. Massa, and S. Piffer

January 2011

Technical Report # DISI-11-278

ON THE USE OF SVM FOR ELECTROMAGNETIC SUBSURFACE SENSING

A. Boni, M. Conci, A. Massa, and S. Piffer

Department of Information and Communication Technologies,
University of Trento, Via Sommarive 14, I-38050 Trento, Italy.
E-mail: andrea.massa@ing.unitn.it; Web-page: http://www.eledia.ing.unitn.it

Abstract

In this paper, a classification approach for the real-time identification of "occupation" areas (instead of the detection of each subsurface object) in sub-surface sensing applications is applied. A suitable SVM-based strategy is developed for determining the probability of occurrence of buried targets and to define a "risk map" of the investigation domain. To assess the effectiveness of the proposed approach and to evaluate its robustness, selected numerical results related to a two-dimensional geometry are presented.

Introduction

In clearing terrains contaminated or potentially contaminated by landmines and/or unexploded ordnances (UXOs), a quick wide-area surveillance is often required. Such a process is inevitably time-expensive and it involves complex acquisition procedures. Consequently, high costs should be met. This is one of the main motivation of the growing research interest in developing unsupervised techniques able to effectively (in terms of time and resources) repair landmine/UXO contaminated areas.

Several solutions have been proposed based on various methodological approaches (see, for instance, [1] and the references cited therein), which consider different sensor modalities such as ground-sensors or synthetic aperture radars. In such a framework, electromagnetic approaches based on learning-by-examples (LBE) techniques [2][3] have been recently proposed for the on-line (after the *training phase*, performed once and off-line) detection of subsurface objects.

However, because of the complexity of the underlying architecture, some difficulties occur when a larger number of unknowns is taken into account. As a consequence, LBE regression-based approaches turn out to be very effective for the detection of a single (or few) buried object, whereas they are not-so-suitable in dealing with the detection of multiple targets. On the other hand, it should be pointed out that the identification of free-areas and the estimation of the concentration of subsurface objects (instead of the localization of each buried scatterer) might be enough in several situations. Then, the goal of a subsurface sensing technique could be moved from the "*object detection*" to the "*definition of a risk map*". Consequently, a classification approach, instead of a regression one, could be employed.

In this paper, such an approach is preliminary investigated and assessed through a numerical analysis with two-dimensional geometries in noiseless as well as noisy conditions.

Mathematical Formulation

Let us consider a half-space subsurface scenario where the upper region presents the same characteristics of the vacuum ($\varepsilon_{r1}=1.0$ and $\sigma_1=0.0$) and the lossy subsurface region is described by ε_{r2} and σ_2. The extension of the investigation domain (where the unknown objects lie) below the surface is $D_I=\{-L/2\le(x,y)\le L/2\}$. T transmitters (z-directed electric line sources) located above the air-ground interface radiate an "incident field" $\underline{E}_{inc}=E_{inc}(x,y)\hat{\underline{z}}$. The "scattered field", $\underline{E}_{scat}=E_{scat}(x,y)\hat{\underline{z}}$, is collected by a set of R sensors close to the air-ground interface.

To define a "*risk map*", let us model the investigation domain with a two-dimensional lattice of M. The state χ_m of the m-th cell can be either empty (if any scatterer belongs to the cell) $\chi_m=-1$ or occupied $\chi_m=1$. Then, the problem is recast as "*to determine the probability array* $\underline{Q}=\{q_m, m=1,...,M\}$" defined as $\underline{Q}=\Pr\{\underline{\chi}=\underline{1}\,|\,\underline{\Gamma}_E\}$, where q_m is the probability that the m-th cell is occupied and $\underline{\Gamma}_E=\{E_{scat}^{(t)}(x_r,y_r); r=1,...,R; t=1,...,T\}$. Such a classification problem can be solved by means of an approach based on a support-vector-machine (SVM) starting from the knowledge of a set of known examples (i.e., input-output relations $\{(\underline{\Gamma}_E, m, \chi_m; m=1,...,M)^{(n)}; n=1,...,N\}$ called training set).

SVM-based Classification Approach

The proposed SVM-based classification approach is formulated as a two-step procedure

Step 1: to determine a decision function $\hat{\Phi}$ that correctly classifies an input pattern $(\underline{\Gamma}_E, m)$ (not-necessarily belonging to the training set);

Step 2: to map the decision function $\hat{\Phi}\{(\underline{\Gamma}_E, m)\}$ into an *a-posteriori* probability $\Pr\{\underline{\chi}=\underline{1}\,|\,\underline{\Gamma}_E\}$

***Step 1:* Definition of the Decision Function**

At this step, the status χ_m of each cell of the lattice has to be determined. Mathematically, such a problem formulates in the definition of a suitable discriminant function $\hat{\Phi}$ that separates the two classes $\chi = 1$ and $\chi = -1$. SVM defines a linear decision function corresponding to a hyperplane that maximizes the separating margin between the classes. Such a linear data-fitting is carried out in the "*feature space*" $\aleph\left\{\underline{\varphi}\left(\underline{\Gamma}_E\right)\right\}$ (different from the original input space $\Re\left\{\underline{\Gamma}_E\right\}$) where the original examples are mapped through a non-linear operator $\underline{\varphi}(\bullet)$. Consequently, the nonlinear SVM classifier is defined as

$$\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E, m\right)\right) = \underline{w} \cdot \underline{\varphi}\left(\underline{\Gamma}_E, m\right) + b,\ m = 1, \ldots, M \qquad (1)$$

where $\underline{w}$ and b are the parameters to be determined during the *training phase*. $\underline{w}$ is a linear combination of the mapped vectors $\underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)$

$$\underline{w} = \sum_{m=1}^{M} \sum_{n=1}^{N} \left\{\alpha_m^{(n)} \chi_m^{(n)} \underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right)\right\} \qquad (2)$$

where $\alpha_m^{(n)} \geq 0$, $n = 1, \ldots, N$, $m = 1 \ldots, M$ are the unknown Lagrange multipliers. Moreover, from the Karush-Khun-Tucker conditions at the optimality [4], b turns out to be expressed as follows

$$b = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N_{sv}} \left\{\chi_m^{(n)} - \sum_{p=1}^{N} \left\{\alpha_m^{(p)} \underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right) \cdot \underline{\varphi}\left(\underline{\Gamma}_E^{(p)}, m\right)\right\}\right\}}{N_{sv}} \qquad (3)$$

N_{sv} being the number of patterns $\left(\underline{\Gamma}_E^{(n)}, m\right)$ for which $\alpha_m^{(n)} \neq 0$ (called "*support vectors*").

In order to determine $\left\{\alpha_m^{(n)}; n = 1, \ldots, N; m = 1 \ldots, M\right\}$, it is necessary the minimization of the following cost function

$$\Omega\left(\underline{w}\right) = \frac{\left\|\underline{w}\right\|^2}{2} + \lambda_+ \sum_{m=1}^{M} \sum_{n=1}^{N_{(m)}^+} \xi_{(m)_+}^{(n)} + \lambda_- \sum_{m=1}^{M} \sum_{n=1}^{N_{(m)}^-} \xi_{(m)_-}^{(n)} \qquad (3)$$

subject to the separability constraints

$$\begin{aligned} &\underline{w} \cdot \underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right) + b \geq 1 - \xi_{(m)_+}^{(n)} \quad \text{for } \chi_m^{(n)} = 1 \ m = 1, \ldots, M \\ &\underline{w} \cdot \underline{\varphi}\left(\underline{\Gamma}_E^{(n)}, m\right) + b \geq \xi_{(m)_-}^{(n)} - 1 \quad \text{for } \chi_m^{(n)} = -1 \ n = 1, \ldots, N \end{aligned} \qquad (4)$$

where $\xi_{(m)}^{(n)}$ are the "*slack variables*"; $N_{(m)}^+$ and $N_{(m)}^-$ indicate the number of training examples for which $\chi_m^{(n)} = 1$ and $\chi_m^{(n)} = -1$, respectively; $\lambda_+ = C \Big/ \sum_{m=1}^{M} N_{(m)}^+$ and $\lambda_- = C \Big/ \sum_{m=1}^{M} N_{(m)}^-$ [5]; the user-defined hyperparameter C controls the trade-off between the empirical risk and the model complexity to avoid the overfitting. The arising constrained optimization problem (3)-(4) is reformulated in a more practical dual form

$$\begin{aligned} &\max_{\underline{\alpha}} \left\{\Omega_{Dual}\left(\underline{\alpha}\right)\right\} = \\ &\max_{\underline{\alpha}} \left\{ \frac{\sum_{n=1}^{N} \sum_{p=1}^{N} \sum_{m=1}^{M} \left[\alpha_m^{(n)} \alpha_m^{(p)} \chi_m^{(n)} \chi_m^{(p)} \Theta\left(\underline{\Gamma}_E^{(n)}, \underline{\Gamma}_E^{(p)}, m\right)\right]}{2} \right. \\ &\left. - \sum_{m=1}^{N} \sum_{m=1}^{M} \alpha_m^{(n)} \right\} \end{aligned} \qquad (5)$$

subject to $\sum_{n=1}^{N} \sum_{m=1}^{M} \alpha_m^{(n)} \chi_m^{(n)} = 0$, $\alpha_m^{(n)} \in \left[0, \lambda_-\right]$ if $\chi_m^{(n)} = -1$ and $\alpha_m^{(n)} \in \left[0, \lambda_+\right]$ otherwise. Since $\Omega_{Dual}\left(\underline{\alpha}\right)$ is a convex and quadratic function of $\alpha_m^{(n)}$, it is solved numerically by means of a standard quadratic programming technique (e.g., the Platt's SMO algorithm for classification [6]). When the Lagrange multipliers and b are computed, then $\hat{\Phi}$ turns out to be

$$\begin{aligned} &\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E\right), m\right) = \sum_{m=1}^{M} \sum_{n=1}^{N} \left\{\alpha_m^{(n)} \chi_m^{(n)} \Theta\left(\underline{\Gamma}_E^{(n)}, \underline{\Gamma}_E, m\right)\right\} + \\ &\frac{\sum_{m=1}^{M} \sum_{n=1}^{N_{sv}} \left\{\chi_m^{(n)} - \sum_{p=1}^{N} \left\{\alpha_m^{(p)} \Theta\left(\underline{\Gamma}_E^{(n)}, \underline{\Gamma}_E^{(p)}, m\right)\right\}\right\}}{N_{sv}} \end{aligned} \qquad (6)$$

where $\Theta\left(\underline{\Gamma}_E^{(i)}, \underline{\Gamma}_E^{(j)}, m\right) = \underline{\varphi}\left(\underline{\Gamma}_E^{(i)}, m\right) \cdot \underline{\varphi}\left(\underline{\Gamma}_E^{(j)}, m\right)$ is a suitable kernel function.

Step 2: ***Mapping of the Decision Function into the A-Posteriori Probability***

Unlike standard SVM classifiers that labels an input pattern according to the following rule [7]

$$\chi_m = sign\left\{\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E, m\right)\right)\right\} \quad m = 1,\ldots,M\ , \qquad (7)$$

the proposed approach is aimed at defining the *a-posteriori* probability $\Pr\left\{\underline{\chi} = \underline{1} \,|\, \underline{\Gamma}_E\right\}$. Towards this end, the *a-posteriori* probability is approximated with a sigmoid function

$$\Pr\left\{\chi_m = 1 \,|\, \left(\underline{\Gamma}_E, m\right)\right\} = \frac{1}{1+\exp\left\{\gamma\hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E, m\right)\right)+\delta\right\}}. \quad (8)$$

$$m = 1,\ldots,M$$

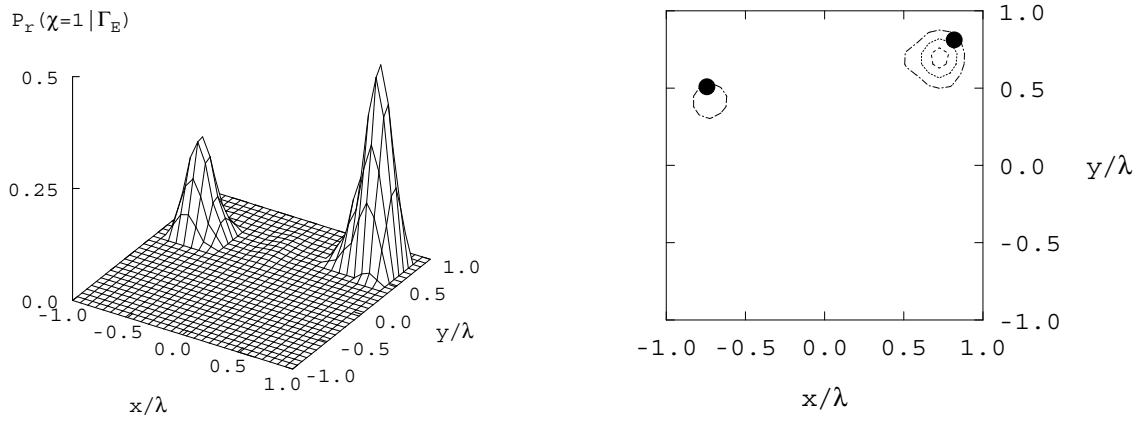
where γ and δ are estimated according to a fitting process. More in detail, a subset of the input patterns of the training set is selected $\left\{\left(\underline{\Gamma}_E, m, \chi_m; m = 1,\ldots,M\right)^{(s)}; s = 1,\ldots,S\right\}$ and the following cost function is defined

$$\Upsilon\left\{\gamma,\delta\right\} = -\sum_{s=1}^{S}\sum_{m=1}^{M}\left\{\frac{\chi_m^{(s)}+1}{2}\log\left[\frac{1}{1+\exp\left(\gamma\hat{\Phi}_m^{(s)}+\delta\right)}\right]+\left(\frac{1-\chi_m^{(s)}}{2}\right)\log\left[\frac{\exp\left(\gamma\hat{\Phi}_m^{(s)}+\delta\right)}{1+\exp\left(\gamma\hat{\Phi}_m^{(s)}+\delta\right)}\right]\right\} \quad (9)$$

where $\hat{\Phi}_m^{(s)} = \hat{\Phi}\left(\underline{\varphi}\left(\underline{\Gamma}_E^{(s)}, m\right)\right)$. Successively (9) is minimized according to the numerical procedure proposed in [8].

Numerical Results

For the experimental validation, the following scenario has been considered. The relative permittivity and the conductivity of the homogeneous subsurface region are $\varepsilon_{r2} = 4.0$ and $\sigma_2 = 10^{-3}$ [S/m], respectively.



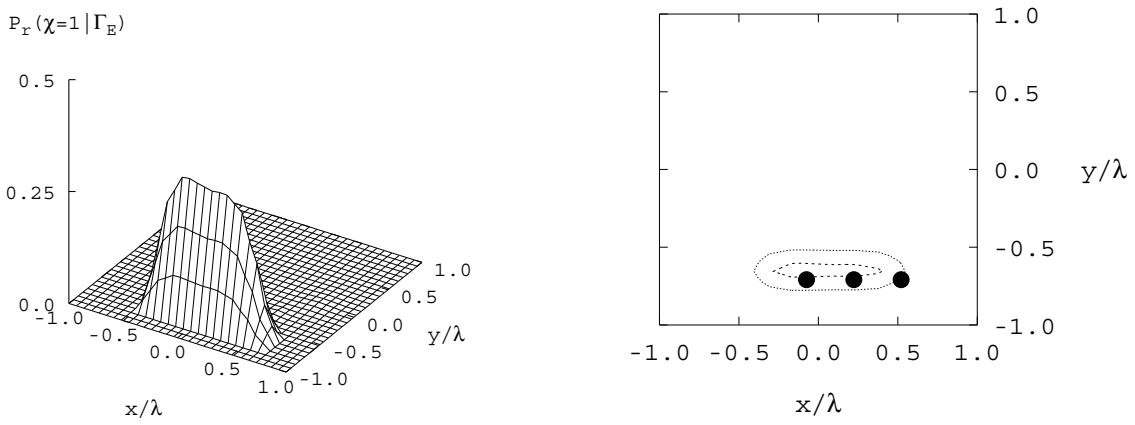
(*a*) (*b*)

Figure 1. Risk map for the two-targets scenario: (a) three-dimensional and (b) contour level representation

The investigation domain is a $2.0\lambda \times 2.0\lambda$ region partitioned in a lattice of $M = 36$ square cells. The buried objects, modeling UXOs or landmines, are lossless circular cylinders of diameter $d = \lambda/6$ with a relative permittivity $\varepsilon_{UXO} = 5.0$. $R = 16$ receivers are equally-spaced along an observation line 2.0λ in length and parallel to the air-ground interface $d = 0.6\lambda$ above the surface. The probing source ($t = 1$) is located at $x_t = 0.0$ and $y_t = 7\lambda/6$. The training is composed of $N = 2484$ patterns related to two and three-targets configurations. These patterns have been also used during the validation test for defining the *a-posteriori* fitting model ($\gamma = -0.533$ and $\delta = 1.272$). Concerning the SVM structure, Gaussian kernel functions were adopted and their parameters selected according to [9].

Within the numerical validation, the first experiment deals with a test set of $P = 2484$ patterns (related to examples different from those of the training phase and concerned with two- and three-scatterers configurations, $P^{(2)} = 1260$ and $P^{(3)} = 1224$) and noiseless conditions. Figs. 1(*a*)-(*b*) and 2(*a*)-(*b*) show the risk maps obtained for two examples of the test set. The first example (Fig. 1) refers to a two-targets configuration where the UXOs are located as indicated in Fig. 1(*b*).

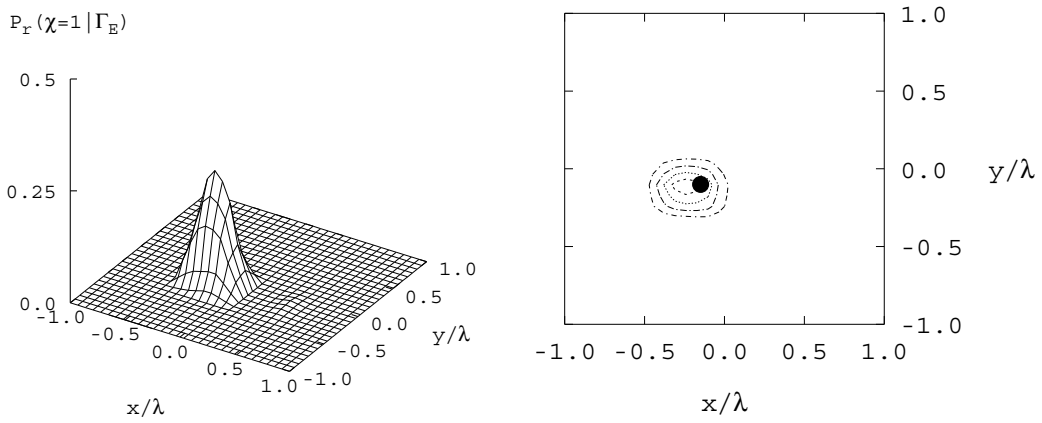
The second example (Fig. 2) is related to a three-scatterers configuration. The objects are adjacent and lie at the bottom of the investigation domain. As expected, when the targets are buried far from the surface, the localization of the "dangerous zones" is more difficult.



(*a*) (*b*)

Figure 2. Risk map for the three-targets scenario: (*a*) three-dimensional and (*b*) contour level representation.

The second numerical experiment considers a more critical scenario where a single target is supposed to be located in the investigation domain ($P^{(1)} = 1296$). It should be pointed out that such a configuration does not belong to the training set. As an example, the risk map for a sample of the test set is shown in Fig. 3.



(*a*) (*b*)

Figure 3. Risk map for the single-target scenario: (*a*) three-dimensional and (*b*) contour level representation.

References

[1] *IEEE Trans. Geosci. Remote Sens.*, Special Issue on: "New Advances in Subsurface Sensing: Systems, Modeling and Signal Processing," vol. 39, Jun. 2001.

[2] S. Caorsi, D. Anguita, E. Bermani, A. Boni, M. Donelli, and A. Massa, "A comparative study of NN and SVM-based electromagnetic inverse scattering approaches to on-line detection of buried objects," *J. Applied Computat. Electromagnetics Soc.*, vol. 18, pp. 1-11, 2003.

[3] E. Bermani, A. Boni, S. Caorsi, and A. Massa, "An innovative real-time technique for buried object detection" *IEEE Trans. Geosci. Remote Sens.*, vol. 41, pp. 927-931, 2003.

[4] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge, University Press, 2000.

[5] K. Morik, P. Brockhausen, and T. Joachims, "Combining statistical learning with a knowledge-based approach: a case study in intensive care monitoring", *Proc. 16th Int. Conf. Machine Learning*, MIT Press., 1999.

[6] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel methods - Support Vector Learning*, B. Scholkopf, C. J. C. Burges, and A. J. Smola (Eds.), MIT Press, 1999.

[7] K. -R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An introduction to kernel-based learning algorithms," *IEEE Trans. Neural Networks*, vol. 12, pp. 181-201, Mar. 2001.

[8] J. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," in *Advances in Large Marging Classifiers*, A. J. Smola, P. Bartlett, B. Scholkopf, D. Schuurmans (Eds.), MIT Press, 1999.

[9] D. Anguita, S. Ridella, F. Rivieccio, and R. Zunino, " Hyperparameter Design Criteria for Support Vector Machines," *Neurocomputing*, vol. 55, pp. 109-134, 2003.