Neurocomputing 168 (2015) 464–474

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A prediction algorithm for data analysis in GPR-based surveys

J.B. Rodriguez [a], M.F. Pantoja [a,*], X.L. Travassos [b], D.A.G. Vieira [c], R.R. Saldanha [d]

[a] Department of Electromagnetismo y Física de la Materia, Universidad de Granada, Granada 18071, Spain
[b] Centro de Engenharias da Mobilidade, Universidade Federal de Santa Catarina, Joinville, Brazil
[c] ENACOM Handcrafted Technologies, Belo Horizonte, Brazil
[d] Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

ARTICLE INFO

Article history:
Received 1 November 2014
Received in revised form
1 April 2015
Accepted 22 May 2015
Communicated by V. Palade
Available online 1 June 2015

Keywords:
Ground-penetrating radar
Neural network applications
Radar signal processing
Geophysics

ABSTRACT

This paper presents a prediction algorithm for features detection in Ground Penetrating Radar (GPR) based surveys. Based on signal processing and soft-computing techniques, the coupled use of principal-component analysis and neural networks enable a definition of an efficient method for analyzing GPR electromagnetic data. To guarantee a low error rate, a study of the algorithm main numerical parameters was performed by means of electromagnetic synthetic-data models. Results for detecting features of geological layers demonstrate not only the method predictions accuracy but also the simple interpretation of its output through scenarios reconstructed images.

1. Introduction

In the last decades, the advent of commercial purposes ground-penetrating radar (GPR) has led to multidisciplinary revolution in the field of buried-object detection, with broad application in areas such as archaeology (e.g., planning of surveys) [1], geology (e.g., aquifer detection) [2], and military industry (e.g., non-metallic mine detection) [3]. One of the main challenges in GPR systems, beyond the mere detection of buried objects, is to gather information on the objects composition or the environment surrounding them.

Although the electronic technology necessary for implementing these systems is now mature with constant developments [4,5], limitations persist in the detection and interpretation of the results provided. In recent years, two main lines have emerged to solve this problem. On the one hand, some systems apply tomographic techniques [6] as well as approaches using integral equations [7], but these have had only partial success due mainly to the field data complexity, which contains high levels of noise caused by non-homogeneities of the host media. On the other hand, techniques based on Neural Networks (NN) with different topologies [8–12] have been proposed to solve canonical electromagnetic-inversion problems, – e.g., a spheroid embedded in a host medium [13], and further improvements have been introduced in relation to more realistic geometrical forms related to civil-engineering applications [14], even including the consideration of a non-homogenous host medium [15].

A common point in all these NN is the implementation, as a step prior to the training phase NN, of a computational model of GPR scenarios. In this way, the scattered field in a randomly generated scenario can be calculated by numerical methods, usually finite differences in the time-domain (FDTD) [16,17] or, for cases where numerical instabilities arise, the alternating-direction implicit FDTD (ADI-FDTD) method [18,19]. One of the main shortcomings of applying NNs as a prediction system in GPR problems is the curse of dimensionality, which makes the training slow and the system prediction capacity poor [20].

Therefore, a key point is to reduce the high dimensionality of the scattered field data, enabling a reduced number of inputs for which the NN will be trained, making the process faster and more reliable. At this point, the application of signal-processing techniques such as Principal-Component Analysis (PCA) can be introduced as part of the algorithm. The usefulness of PCA as a compression technique with minimum loss of information in time-domain GPR signals has been shown in [15]. In this previous work, the objective is to estimate the depth and radius of buried tubes in a non-homogenous concrete structure.

In this context, the present paper seeks to apply techniques based on PCA and NNs to build prediction systems for geological features in GPR-based geological surveys. Furthermore, this procedure main challenge is not only to achieve a high-rate of success in the predictions but also to build on previous works in this research line by producing B-scan graphic results. In this sense, the proposed algorithm outperforms previous NNs predictors, which

* Corresponding author.
E-mail address: mario@ugr.es (M.F. Pantoja).

http://dx.doi.org/10.1016/j.neucom.2015.05.081
0925-2312/

provide one-dimensional numerical outputs, enabling the interpretation of the solutions by users not specialized in the GPR data processing.

The paper is structured as follows. First, a general overview of the background theory is provided, briefly describing the PCA algorithm, the NNs and the synthetic data creation with FDTD. Nextly, the scheme of the prediction system is presented, paying special attention to the differences in the implementation for A- and B-scan surveys. Then, another section shows the influence of some numerical parameters in the performance of the prediction system and, finally, illustrative examples related to the detection and prediction of geological layers are provided.

2. Background theory

Fig. 1 shows the flowchart synthesizing the prediction system. The prediction algorithm can be described as a modular system which combines three different resources in the process: (1) numerical electromagnetic simulation codes, (2) signal-processing compression techniques, and (3) neural-networks theory. Further improvements and incoming advances in any of these theories could be accommodated separately at each stage of the process.

The first step in developing a NN-based prediction algorithm is to gather representative situations data in which the neural network will work. Successful accomplishment of the NN training and configuration phases will be directly related to the diversity, quantity, and quality of the data provided. For GPR systems, the use of experimental data is hardly affordable because (1) it is time consuming and labor intensive, and (2) it is scarcely free of undesired objects and other experimental sources of errors. For these reasons, the use of incoming data from electromagnetic simulations is considered.

The FDTD numerical approach for the solutions of Maxwell's equations has been broadly employed for GPR simulations [17]. These provide higher accuracy than ray-tracing methods [21] at a cost of increasing the computational burden. Moreover, realistic GPR scenarios can be solved due to the ability to deal with non-homogeneous and dispersive materials. However, in some cases, numerical instabilities can arise, invalidating the computed results. In such a case, an improved numerical version of the FDTD, called ADI-FDTD, which is based on an implicit finite-difference formulation of the Maxwell time-domain equations, can generate accurate results. Therefore, proper modeling of the GPR equipment and different possible scenarios (e.g., electromagnetic sources, feeding pulse, constitutive materials parameters, geometries of non-homogeneous soils) can be efficiently introduced and solved with the aid of scripts. The process automation is required since the training phase typically needs to run hundreds of cases until the NN can be determined.

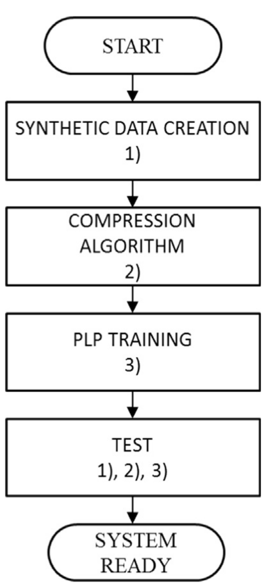

Fig. 1. Flowchart of the prediction system. Numbers mean for external resources applied at each step (see text for details).

Signal-processing compression techniques have constituted an active field of research in the last few decades, mainly for applications related to audio and image processing [22]. Designed initially for communication systems, they are aimed at handling a large amount of information with the least data possible. In this sense, the problem considered here is analogous. The huge amount of data obtained from GPR electromagnetic simulations makes their direct use inefficient for the NN configuration, mainly due to the high complexity of the training algorithms, which require the handling of incoming data from hundreds of simulations in order to determine the variables and NN weights. Even in scenarios where a sufficiently high number of simulations can be calculated, it is possible that advanced training algorithms does not converge for high-dimensional NNs, primarily due to the difficulty of providing a non-sparse set of training data [23,24].

For this reason, it becomes necessary to process the synthetic data and remove the redundant information. This redundancy of data is a typical feature in the GPR systems, where exhaustive measurements are made over the same scenario and minor differences between adjacent traces appear. To exploit the strong correlation in the data, Principal-Components Analysis (PCA) can be applied [25]. PCA identifies similar patterns in data, and reorganizes the data in such a way that the similarities and differences are highlighted. Mathematically, this is achieved by an orthogonalization of a matrix constructed by adding rows with traces of input data, so that these rows are not correlated with each other. Another main feature of PCA is that once these patterns in the data are found (e.g., the orthogonal basis is determined), the data can also be compressed without significant loss of information by simply removing some of the basis vectors.

In the present paper, the compression ability of PCA is used to extract the most relevant information from the data set employed in the NN training phase. Naming the principal components of a given trace as its components in the new orthogonal basis derived from the initial matrix data, composed of a large amount of examples, the NN input is precisely these principal components. As a reduction in the basis dimension can be performed with a low impact on the original trace, the number of principal components can be reduced leading thus to low-dimensional NNs. The next section will show that the decomposition of different GPR signals using the PCA is a key factor of the NN ability to reconstruct the original geological scenarios.

Finally, some considerations concerning the NNs are necessary to finish the description of the theories on which the proposed system is founded. The initial development of the theoretical basis of NN systems [26–28] has in recent years been followed by a mature period of real-world applications [29–31]. A major decision in any problem to be solved is the choice of the network topology, because this topology has a significant impact on the system performance, and the same problem can often be successfully solved with different topologies. Among a considerable number of network topologies [32,33], the work described here has used a particular topology called Parallel-Layer Perceptron (PLP) [24,10]. Regarding the inversion problem, this topology has some advantages over other classical networks, such as the multilayer perceptron (MLP) [33] and the adaptive-network-based fuzzy-inference system (ANFIS) [34].

In addition, PLP offers better performance than MLP while maintaining the ability of ANFIS to handle complex problems. The training algorithm for PLP used in the present paper is a hybrid method based