Random Forest Regressor for Layered Earth data inversion

Abhinav Pratap Singh[1], Divakar Vashisth[1], Shalivahan Srivastava[1]

(abhinav.17je003487@agp.iitism.ac.in, divakar.16je002079@agp.ism.ac.in)

[1]Indian Institute of Technology (Indian School of Mines), Dhanbad

AGU 100 FALL MEETING San Francisco, CA | 9–13 December 2019

Abstract

Interpretation of geophysical data sets involves solving an inverse problem for recovering information on subsurface physical properties from observed data. Geophysical inversion problems are ill-posed non-linear problems. We have attempted to achieve global convergence for this problem using a Machine learning approach- Random Forest Regressor (RFR). It fits a number of decision trees on various equal sub-samples of dataset and averages the result from each decision tree to give an output. RFR approach has been preferred due to its ability to solve highly complex inversion problems, if trained using sufficient number of training examples without the need of computing forward code in each iteration to determine the output. In this study, inversion of synthetic magnetotelluric (MT) and DC resistivity data sets for a three-layered earth has been performed.

For comparison sake, MT results have been compared to those obtained from Particle Swarm Optimization (PSO), Genetic algorithm (GA) and ridge regression (RR) algorithm. The resistivity results have been compared with PSO and Grey Wolf optimization (GWO) techniques.

The results obtained for MT apparent resistivity problem from RFR are in close agreement with true model parameters and are at par or better than those obtained from RR, GA and PSO.

The considered model for resistivity is a typical equivalence problem. The obtained results are closer to the true model and are at par or better than those obtained from GWO and PSO.

The obtained results demonstrates the efficacy of the regressor especially in tackling equivalence problem.

Introduction

Interpretation of geophysical data sets involves solving an inverse problem for recovering information on subsurface physical properties from observed data. Geophysical inversion problems are ill-posed non-linear problems. The solutions to such a problem are estimated by using the global optimization techniques like swarm intelligence techniques through the computation of forward code in each step.

In the present work, we assume the 1-D earth model and each layer is electrically homogenous and isotropic (Cagniard, 1953). For this model, inversion of 1-D MT data and DC resistivity data has been worked out using a complex Machine Learning approach- Random Forest Regressor (RFR).

RFR has the universal approximation property, which means that they can approximate any function in any dimension and up to a desired degree of accuracy. This property of RFRs is brought into application in the inversion of the MT and DC resistivity data. The resistivity and thickness can be assumed to form a multi-dimensional curve with the resistivity profile values. The aim of the model is to fit this multi-dimensional curve and thus, estimating the resistivity and thickness from the resistivity profile.

Theory

Random forests are a type of ensemble method which makes predictions by averaging over the predictions of several independent base models called decision trees. Since its introduction by Breiman (2001) the random forests framework has been extremely successful as a general purpose classification and regression method.

Random Forest Regressor works by dividing the training dataset equally and randomly into a certain number of groups. Each dataset is then given as input to a decision tree which then processes the dataset and gives its own output. The RFR takes into account the prediction from all the trees and then arrives to a result, generally obtained by averaging out the predictions from all the decision trees.

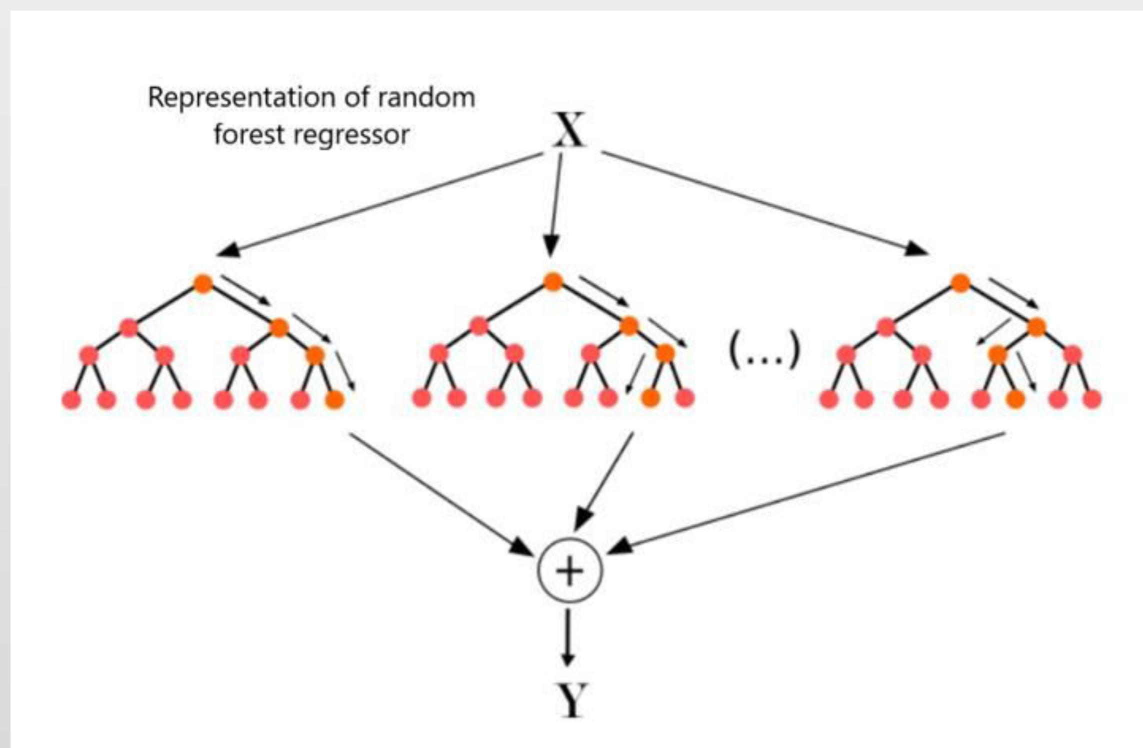

***Figure 1**: A schematic random forest regressor*

Results

Magnetotelluric (MT) dataset

To assess the applicability of RFR in inverting geophysical datasets, we have taken a synthetic MT resistivity profile for a 3-layered horizontally structured earth to train the RFR. The synthetic MT example has been taken from Shaw and Srivastava, (2007). The values obtained after the inversion from various techniques are summarised in Table 1. The MT apparent resistivity profile for true and inverted model parameters obtained from RFR and PSO inversion has been demonstrated in Figure 2.

DC resistivity dataset

We considered the three-layered earth data used by Dixon and Doherty (1977) to compute apparent resistivity data for a Schlumberger sounding experiment. Table 2 shows the values of the search spaces and model parameters computed from RR, GA, PSO and RFR inversion.

Figure 3 shows the DC apparent resistivity profile for true model parameters and the ones obtained from RFR and PSO inversion.

Parameter	True values	RR (Shaw and Shalivahan, 2007)	Search Space		GA (Shaw and Shalivahan, 2007)	PSO (Shaw and Shalivahan, 2007)	RFR
			Minimum	Maximum			
ρ_1 (Ωm)	30,000	43,424.4	5,000	50,000	40,800	26,981.8	28,281.9
ρ_2(Ωm)	5,000	3,097.1	1,000	10,000	10,000	6,230.3	4,958.6
ρ_3(Ωm)	1,000	980.7	50	5,000	1,010	1,011.7	998.8
h_1(km)	15.0	17.01	5.0	25.0	6.21	13.09	16.8
h_2(km)	18.0	16.96	10.0	25.0	25.0	19.72	19.9

***Table 1**: True parameters, search space, and results of inversions using RR, GA, PSO, and RFR for a synthetic MT apparent resistivity dataset.*

Parameter	True values	RR (Shaw and Shalivahan, 2007)	Search Space		GA (Shaw and Shalivahan, 2007)	PSO (Shaw and Shalivahan, 2007)	RFR
			Minimum	Maximum			
ρ_1 (Ωm)	2,500	2,300	1,000	5,000	2,760	2784.9	2,123.72
ρ_2(Ωm)	100	90	10	1,000	121	110.1	110.1
ρ_3(Ωm)	300	270	10	3,000	275	326.5	276.9
h_1(m)	1.5	1.3	0.5	10.0	1.5	1.5	1.47
h_2(m)	25.0	20.0	5.0	50.0	18.0	28.6	24.7

***Table 2**: True parameters, search space, and results of inversions using RR, GA, PSO, and RFR for a synthetic DC resistivity dataset.*

The above table clearly shows that the values obtained from RFR are quite close to the true values. The performance of the RFR is found to be at par or better than other inversion algorithms. The MT apparent resistivity profile from the RFR matches closely with that of the true values, as is visible in Figure 2 and 3 respectively.

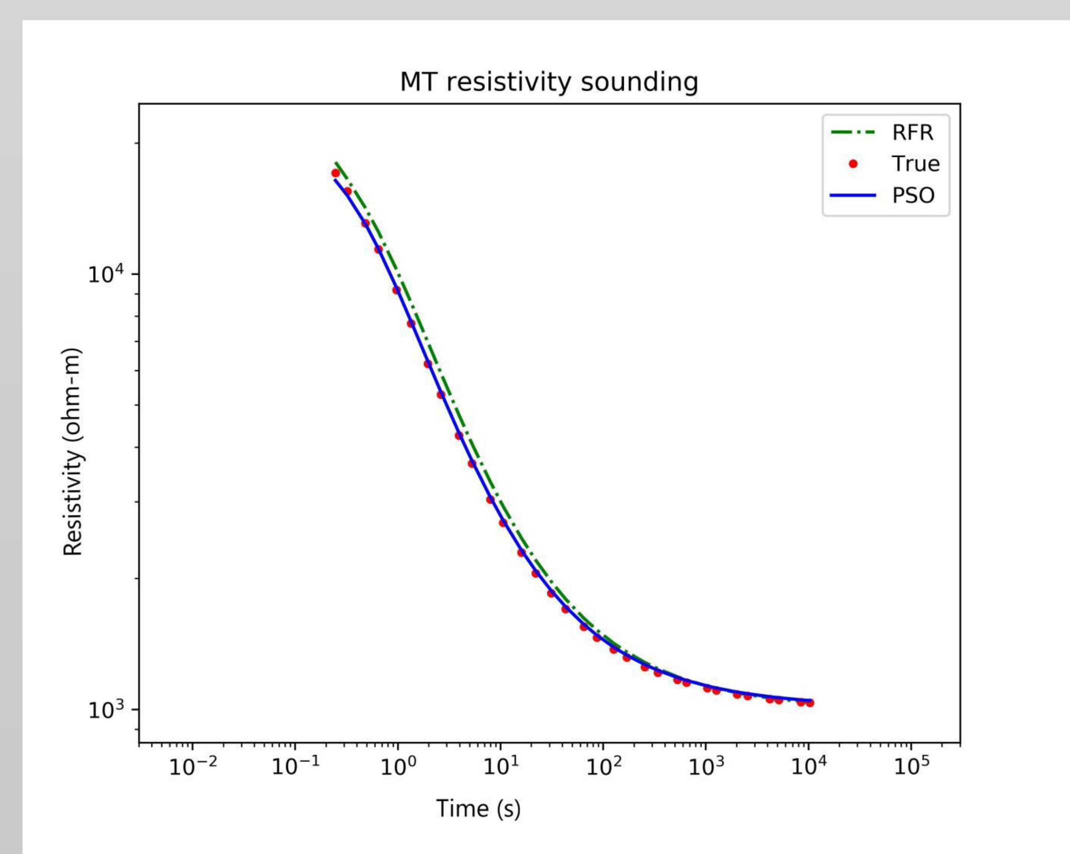

***Figure 2:** MT apparent resistivity profile for true and predicted values using PSO, GA, and RFR*

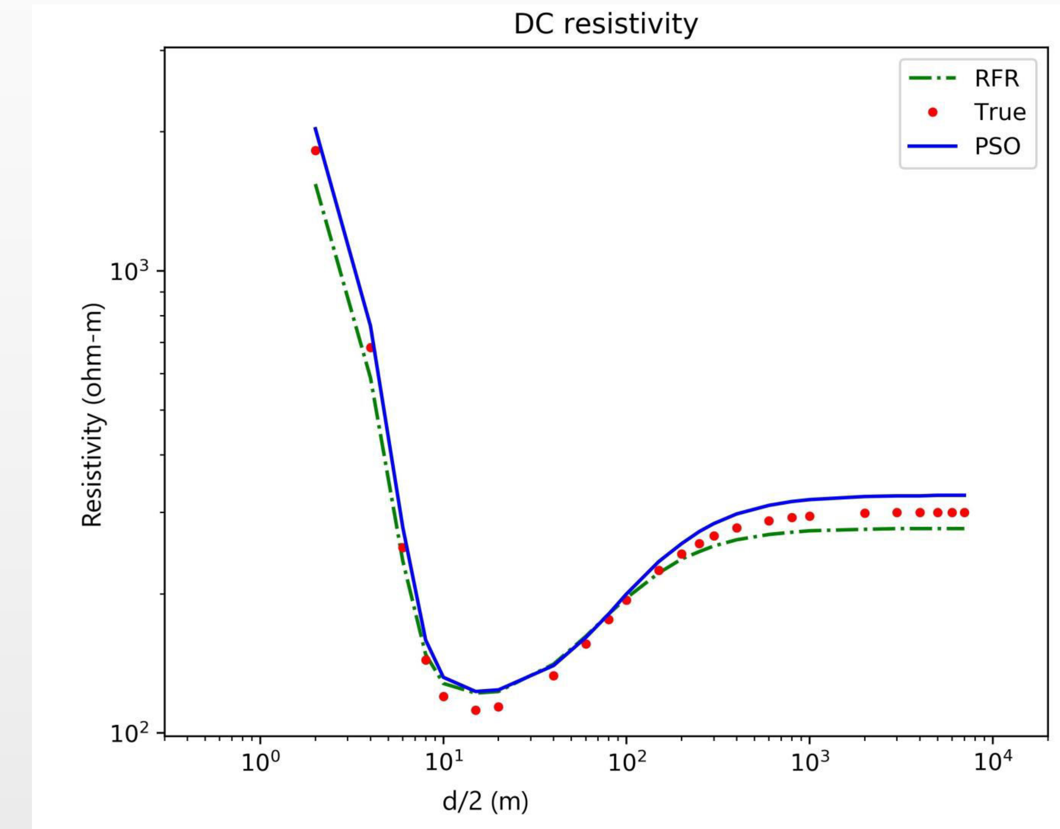

***Figure 3:** DC apparent resistivity profile for true and predicted values using PSO, GA, and RFR*

Conclusion

In this study, we report the utility of the random forest regressor in the geophysical inversion of the 1-D MT and DC resistivity dataset. The RFR is very efficient, robust and simple to be used for inversion. It provides an alternative approach for the inversion of the 1-D MT and DC resistivity data. The algorithm is able to predict model parameters that are in close agreement with the true parameters for the MT apparent resistivity profile and DC resistivity profile. The efficacy of the model is demonstrated through the comparisons with the other well known algorithms.

References

Cagniard, L., 1953, Basic Theory of magnetotelluric method: Geophysic, 18, 605-635.

Dixon, O., and J. E. Doherty, 1977,New interpretation methods for IP soundings: ASEG Bulletin, 8, 65–69.

Jupp. D.L.B., and K. Vozoff, 1975, Stable iterative methods for the inversion of geophysical data : Geophysical Journal of the Royal Astronomical Society, 42, 957-976.

L Breiman , 2001, Random Forests, Machine Learning, 45, 5-32.

Rodi, W., and R.L. Mackie, 2001, Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion: Geophysics, 66, 174-187

Shaw, R.K and S. Srivastava, 2007, Particle swarm optimization: A new tool to invert geophysical data, Geophysics, 72, F75-F83.