What Can Machine Learning Do for Geomagnetic Data Processing? A Reconstruction Application

Poster · May 2018

4 authors, including:

Huan Liu
China University of Geosciences

Zheng Liu
University of British Columbia - Okanagan

Hengli Song
University of Technology Sydney

What Can Machine Learning Do for Geomagnetic Data Processing? A Reconstruction Application

Huan Liu[1], Haobin Dong[1, *], Zheng Liu[2], Hengli Song[1]

1. School of Automation, China University of Geosciences, Wuhan, Hubei, China

2. School of Engineering, University of British Columbia Okanagan Campus, Kelowna, BC, Canada

*Corresponding author email: donghb@cug.edu.cn

Introduction

The integrity of geomagnetic data is a critical factor for understanding the evolutionary process of Earth's magnetic field, as it can provide useful information for near-surface exploration, unexploded explosive ordnance (UXO) detection, etc. **Aimed to reconstruct geomagnetic data from under-sampled or missing traces, this paper presented an approach based on machine learning techniques to avoid the time & labor-intensive nature of the traditional manual and linear interpolation approaches**. In this study, three classic machine learning models, support vector machine (SVM), random forests and gradient boosting were built. The proposed learning models were first used to specify a continuous regression hyperplane from training data, to recognize the probably intrinsic relation between missing and completed traces. Afterwards, the trained models were used to reconstruct the missing geomagnetic traces for validation, while testing other new field data. Finally, numerical experiments were derived. **The results showed that the performance of our methods was more competitive in comparison with the traditional linear method, as the reconstruction accuracy was increased by approximately 10% ~ 15%**.

Methodology

In our approach, we propose to solve the reconstruction problem using machine learning techniques. Moreover, to our best knowledge, we are the first to adopt this method to reconstruct the under-sampled geomagnetic data with missing traces. The reconstruction of geomagnetic data can be modeled as a regression problem. In this study, three classic machine learning models were built for this regression problem, i.e. *Support Vector Machine* (SVM), *Gradient Boosting* and *Random Forests*.

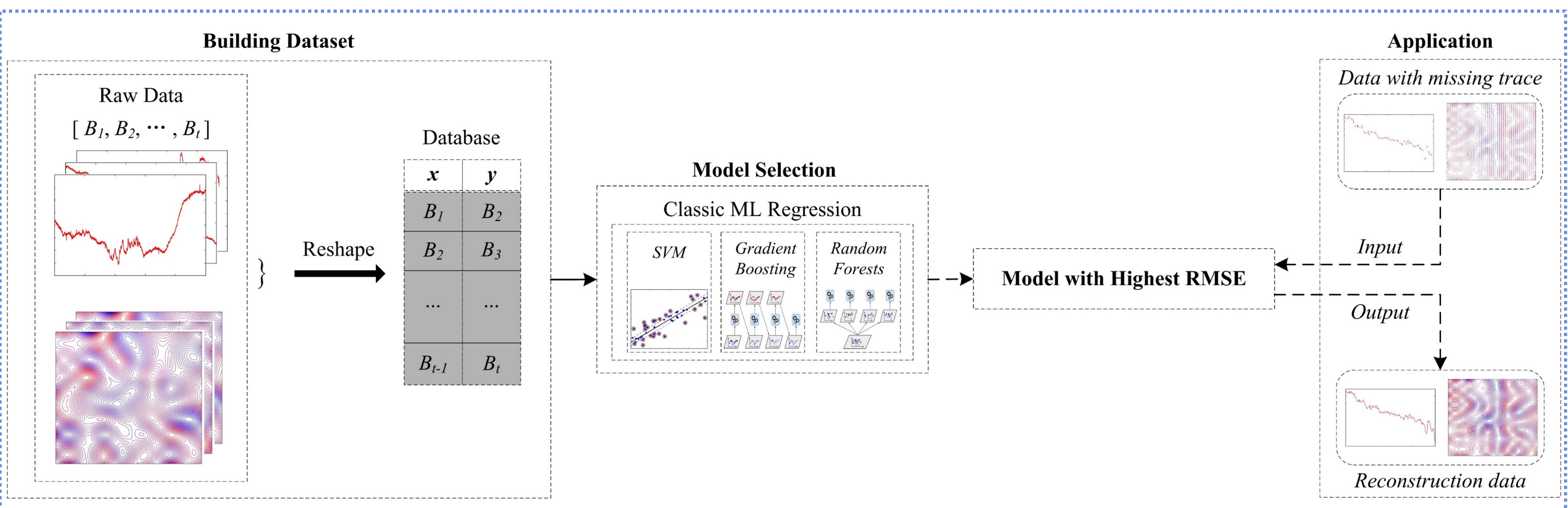


Fig. 1. Pipeline of proposed geomagnetic data reconstruction framework, which includes three main modules: 1) dataset building, 2) classic machine learning regression, and 3) the selected model for regression and reconstruction.

The main steps in our reconstruction method using classic machine learning techniques are given in **Alg. 1**. Following the procedures above, we fed all point pairs from the experimental training data into the each of the regression models to train. Then, three continuous regression models (hyperplane, $f(x)$) can be generated, and it also could be saved for future use in the prediction stage. In the prediction stage, the values of missing geomagnetic data were unknown, but they could be predicted using the values that are before and after them in a period of time. All variables (x) with missing traces were input simultaneously into the regression models trained in the previous stage, thus allowing us to obtain the missing geomagnetic data.

Algorithm 1: Reconstruction based on classic machine learning

Result: Geomagnetic data reconstruction.
Input : Experimental geomagnetic data x, missing geomagnetic data y.
Output: Predicted geomagnetic data.

1. **Training:**
2. A dataset is built where x stand for the strength of magnetic field at time (t) and y stands for the strength of magnetic field at time (t + 1);
3. Building different independent regression models:
4. $clf1 = svr()$
5. $clf2 = RandomForestRegression()$
6. $clf3 = GradientBoostingRegression()$
7. Input all the pairs of (x_t, x_{t+1}) to each of the models to train.
8. **Testing:**
9. The trained models are applied to the geomagnetic data x with missing traces and obtain the predicted missing data y';
10. Evaluating the results to obtain the model with the best RMSE.

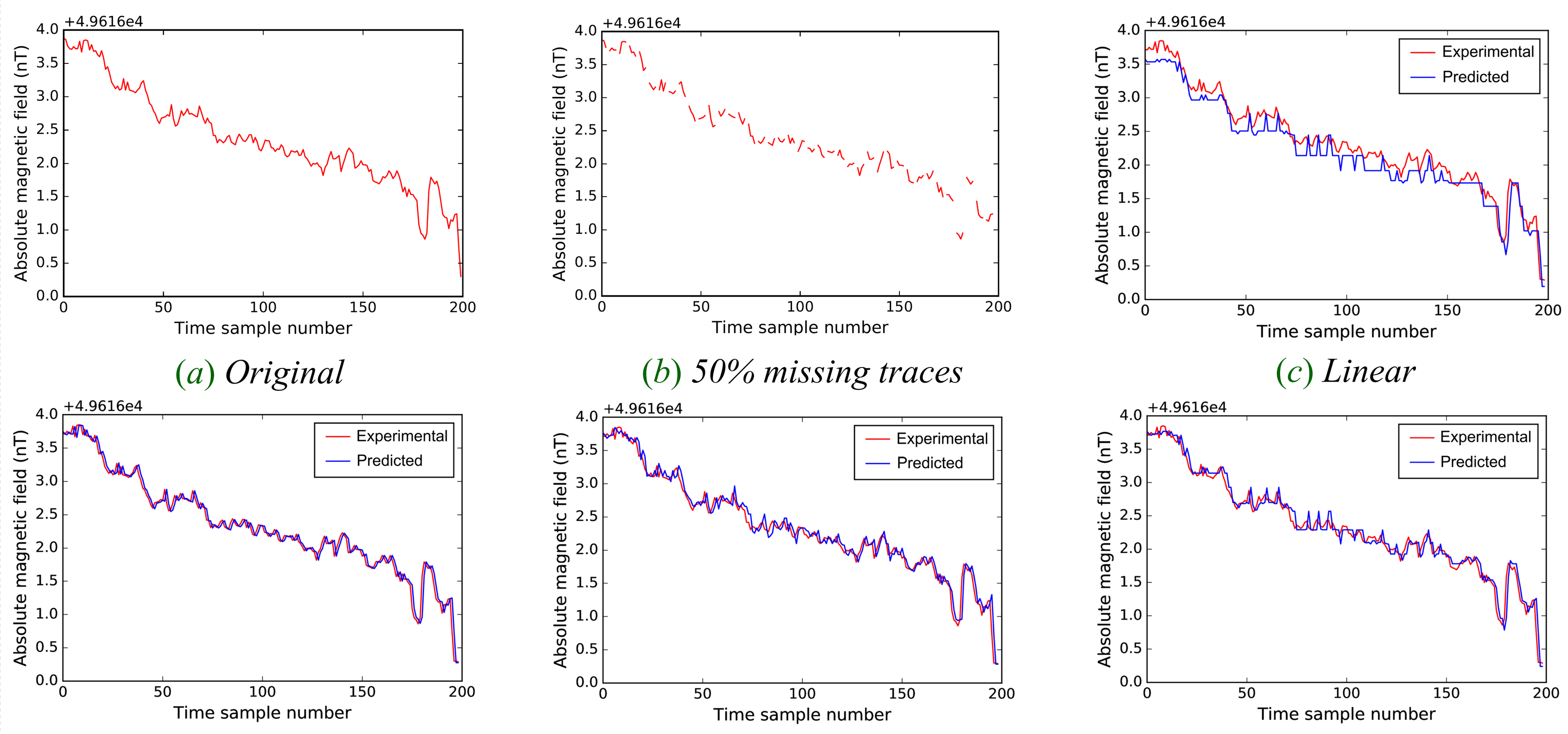
The RMSE and R-squared were used to evaluate the prediction performance as:

$$\begin{cases} RMSE = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \\ R\_squared = 1 - \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \end{cases}$$

y_i : practical value
$\hat{y}_i$: predicted value
$\bar{y}_i$: average of all predictions

Results

We implemented comparisons of the proposed method with the commonly used linear regression method in a case of 50% regular missing traces in 2D and 3D geomagnetic data as shown in **Fig. 2(b)** and **Fig. 3(b)**, respectively. Meanwhile, numerous geomagnetic datasets without missing traces, from which 29728 (2D) and 60000 (3D) training point pairs of feature vectors and labels can be extracted.

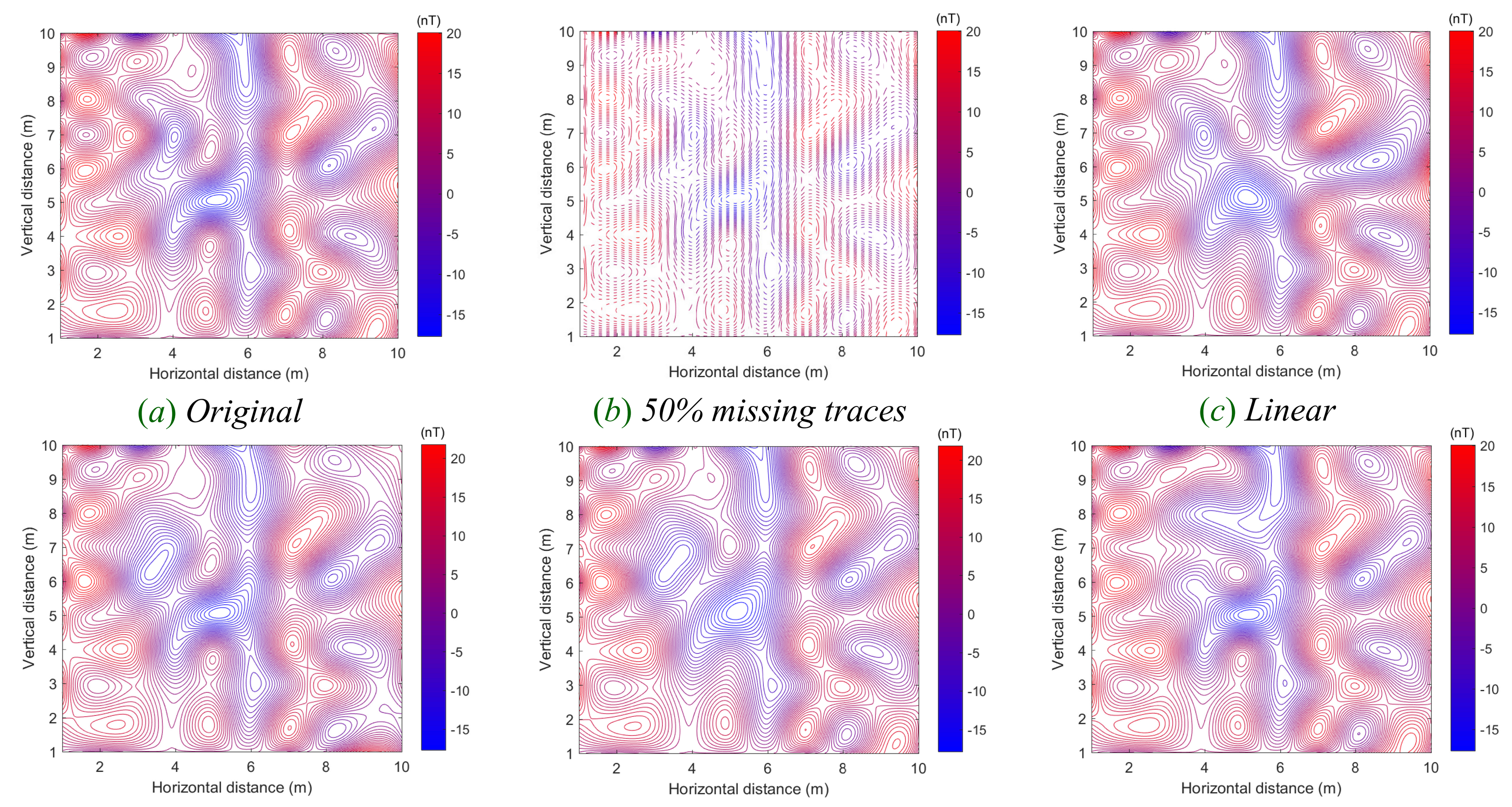


(*a*) *Original* (*b*) *50% missing traces* (*c*) *Linear*
(*d*) *SVM* (*e*) *Random Forests* (*f*) *Gradient Boosting*

Fig. 2. 2D geomagnetic data reconstruction.

Model	dataset	R-squared	RMSE
Linear	Training	**0.811**	**0.253**
	Validation	**0.796**	**0.266**
	Testing	**0.788**	**0.279**
SVM	Training	**0.899**	0.188
	Validation	**0.889**	**0.198**
	Testing	**0.878**	**0.208**
Random Forests	Training	0.896	**0.179**
	Validation	0.885	0.200
	Testing	0.873	0.220
Gradient Boosting	Training	0.893	0.185
	Validation	0.883	0.204
	Testing	0.872	0.223

SVM almost got the best quality of training, validation and testing datasets. The calculated values of R-square and RMSE in training phase were **0.899** and **0.188**, respectively, while the testing results are worse than training. However, when compared to linear, SVM, etc., it demonstrated a significant improvement.



(*a*) *Original* (*b*) *50% missing traces* (*c*) *Linear*
(*d*) *SVM* (*e*) *Random Forests* (*f*) *Gradient Boosting*

Fig. 3. 3D geomagnetic data reconstruction.

Model	dataset	R-squared	RMSE
Linear	Training	**0.795**	**0.269**
	Validation	**0.786**	**0.276**
	Testing	**0.778**	**0.283**
SVM	Training	**0.903**	**0.179**
	Validation	**0.895**	**0.190**
	Testing	**0.887**	**0.201**
Random Forests	Training	0.883	0.192
	Validation	0.876	0.212
	Testing	0.869	0.232
Gradient Boosting	Training	0.879	0.213
	Validation	0.869	0.219
	Testing	0.858	0.225

Our proposed machine learning based methods were also better than the linear regression method as a whole, especially in the area around **coordinate (7, 6)**. Likewise, SVM got the best quality with training, validation as well as testing datasets, which were consistent with the results of 2D example.

Conclusions

In this paper, we present the classic machine learning method to improve the performance of traditional methods. Besides, SVM based approach allowed us to avoid previous drawbacks in existing reconstruction methods, and it is universally applicable to varying datasets. Furthermore, the trained regression model can be saved for future use to reconstruct the geomagnetic data with similar geomorphological structure. **Overall, the experimental results showed that the proposed method can achieve a reconstruction accuracy about 90%, which showed an increased by about 20% than the traditional method.**