International Journal for Modern Trends in Science and Technology, 6(10): 113-117, 2020

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST061020
Available online at: http://www.ijmtst.com/vol6issue10.html

Reconstruction Process of Geomagnetic Data using Machine Learning

D. Venkat Sai[1] | J. Prasanna kumar[2] | Dr. A.V.Krishna Prasad[3]

[1]PG Scholar, Department of CSE, MVSR College of Engineering, Nadergul, Hyderabad.
[2]Professor, Department of CSE, MVSR College of Engineering, Nadergul, Hyderabad.
[3]Associate Professor, Department of CSE, MVSR College of Engineering, Nadergul, Hyderabad.

To Cite this Article
D. Venkat Sai, J. Prasanna kumar and Dr. A.V.Krishna Prasad, "Reconstruction Process of Geomagnetic Data using Machine Learning", *International Journal for Modern Trends in Science and Technology*, 6(10): 113-117, 2020.

Article Info
Received on 30-September-2020, Revised on 12-October-2020, Accepted on 15-October-2020, Published on 23-October-2020.

ABSTRACT

The geomagnetic data plays a important role in understanding the evolutionary process of Earth's magnetic field, as it provides necessary information for near-surface exploration, unexploded explosive ordnance detection, and so on. To reconstruct the geomagnetic data, this project presents a geomagnetic data reconstruction method based on machine learning techniques. The traditional linear approaches are prone to time inefficiency and involves high labor cost, while the proposed approach has a significant improvement. In this project, three classic machine learning models, support vector machine, random forests, and gradient boosting were built. And, a deep learning algorithm, recurrent neural network, was explored to further improve the performance. The proposed learning methods were used to specify a continuous regression hyperplane from a training data. The specified regression hyperplane is a mapping of the relation between the missing data and the surrounding intact data. Then, the trained method, were used to build the missing geomagnetic data for validation, and they can be used for reconstructing further collected new field data. Finally, numerical experiments were derived. The results shows that the performance of our proposed methods was more accurate in comparison with the traditional linear learning method, as the reconstruction accuracy was increased by approximately 10%~20%.

KEYWORDS : *Geomagnetic, Machine Learning, Reconstruction, Support Vector Regression*

INTRODUCTION

The geomagnetic data are carried out with the times spanning from seconds to decades at magnetic observations. However, the integrity of geomagnetic data cannot be guaranteed especially when malfunctions of the observing system happen during the observations. In this case in completed data or missing data could compromise the interpretation accuracy of the geomagnetic data, which necessitate the research on the reconstruction of the geomagnetic data. Till now, several geomagnetic data reconstruction methods were developed. In terms of numerical simulations studies, and investigated how to predict the unknown geomagnetic data based on the data assimilation technique.

A vast array of studies has established the primary tools of machine learning, such as linear regression, decision trees, support vector machine (SVM), artificial neural networks, and instance-based learning. The primary methods performed by machine learning include regression, classification, clustering, and so on. To sum up, according to its good performance, machine learning spreads rapidly throughout multiple fields, suggesting that it is likely to lead the next wave of innovation in geophysics.

The proposed method mainly consists of three steps. The initial step is the process of building training data set. The training data set does not contain missing data, such that the model later trained by this data set is a mapping of the previous values, $x(t)$, to the following values, $x(t+1)$. The data set is randomly classified into three parts: training, validation, and testing data. Next step is to train a regression model(classic machine learning and RNN), which fits a hyperplane from the training data. The final step is to reconstruct the missing geomagnetic data of the mock-up defected data set. Afterward, the reconstructed data are validated and tested, and the best optimistic method is obtained according to the accuracy. The proposed machine learning method only depends on the characteristics of the originally collected data, which can overcome some drawbacks, e.g., the previously mentioned restriction of linear events, the sparsity of reconstruction data, and so on. In consequence, it not only breaks the previous limitations but also shows stronger adaptability in different kinds of the data set. which can help researchers to obtain the complete information of Earth's magnetic field for near-surface exploration and detection for magnetic materials.

Earth's magnetic field, which is also called as the geomagnetic field extends from the Earth's interior out into space, then it interacts with the solar wind, a stream of charged particles emanating from the Sun. This magnetic field is caused by electric currents due to the moment of convection currents of molten iron in the Earth's outer core . these convection currents are caused by heat escaping from the core, a natural process called a geo dynamo. The magnitude of the Earth in its surface ranges from 25 to 65 microteslas (0.25 to 0.65 gauss).As an approximation, which is represented by a field of a magnetic dipole currently tilted .at an angle of about 11 degrees with respect to Earth's rotational axis, as if there is an enormous bar magnet placed at that angle through the center of the Earth. The North geomagnetic pole, which was discovered in 2015 located at Ellesmere Island, Nunavut, Canada, in the northern hemisphere, is actually the south pole of the Earth's magnetic field, and conversely[4].

The geomagnetic field changes with time scales from milliseconds to millions of years. some changes can be traced to geomagnetic storms or daily variations in the currents. Changes over time scales over a year or more mostly reflect changes in the Earth's interior, particularly in the iron-rich core. These changes are referred as secular variation. Secular variation can be observed in measurements at magnetic observatories, where some of them have been around for hundreds of years (the Kew Observatory, for example). Over a period scale, magnetic declination is observed to vary over tens of degrees. A movie on the right shows how global declinations have changed from time to time over the last few centuries [6].

To investigate the longevity and robustness of high-latitude flux patches in the geomagnetic field at the core–mantle boundary, we present time-dependent models of the geomagnetic field for the past 7000 years. Our methods use the same data set as previously used for time-dependent archaeo magnetic field modelling, but constrained with additional priori models from time averages of field models covering the last 150, 400 and 3000 years. We find that the data are consistent with flux patches with both north and south hemispheres for the past 7000 years, and the northern hemisphere patches at least have highly dynamic behaviour. Our results should inform geo dynamo studies of thermal core–mantle coupling, and of possible long-term structure in the geomagnetic field [9].

METHODOLOGY

Geomagnetic Data contains earth magnetic field data and this data will be recorded by sensors and by using this data scientists can know the status of the earth such as when explosion will happen in exploded areas such as volcanoes. Sensors will be configure to read earth magnetic field data based on time intervals such as every minute, seconds or hours. Sometime sensor will miss reporting some data and that missing data can cause serious issues such as missing volcano eruption information. To overcome from such issues various techniques were introduce but those techniques require heavy man power and it's a time consuming task also.
To overcome from this problem here we are using machine learning algorithms to get target information by giving missing values. In this project we are evaluating performance of various

machine learning algorithms such as Support Vector Regression, Random Forest Regression, Gradient Boosting Regression and Deep Learning LSTM Algorithm. In all algorithms LSTM is giving less prediction error compare to other algorithms.

The official Python introduction is

Python is an easy to learn, programming language. It has structured high-level data structures and a simple effective approach to object-oriented programming. It's distinguished syntax and effective typing, together with its explicate nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

Support Vector Regression Algorithm

Support Vector Regression: Support Vector Regression(SVR) is quite different than other Regression data models. It uses Support Vector Machine(SVM, a classification algorithm) algorithm to anticipate a continuous variable. Support Vector Regression attempts to fit the best line within a preset or threshold error value. Then SVR attempts to classify all the speculate lines in two types, ones that pass through the error boundary (space separated by two parallel lines) and ones that don't. Those lines which does not pass the boundary are not considered as the difference between the predicted value and the actual value has exceeded the error threshold. The lines that pass, are considered for a potential support vector to predict the value of an unknown missingvalues.

Random Forest Regression Algorithm

A random forest regression is an combination of randomized regression trees. Denote the predicted value at point by the -th tree, where are independent random variables, distributed as a generic random variable , independent of the sample .Random forest is a bagging technique and not a boosting technique. The trees in random forests runs in parallel. There is no interaction among these trees while building the trees. It runs by setting up a mean of decision trees during training time and producing the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees .A random forest is a meta-estimator which means it combines the result of multiple predictions and aggregates many decision trees, with some helpful modifications.

Gradient Boosting Regression Algorithm:

Gradient boosting is a machine learning concept used for reconstructing data and other classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. Gradient boosting setup the model in a step-by-step fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Deep Learning LSTM(Long Short Term Memory):

Long short-term memory is an pretended recurrent neural network model in the field of deep learning. Unlike other standard feed forward neural networks, LSTM has feedback connections. It not only process single data points, but also whole sequences of data. For instance, LSTM is applicable to tasks such as not divided, connected and involved in handwriting, speech anomaly detection in network traffic or IDSs (intrusion detection systems).

LSTM networks are well capable of categorize,operatingdata and makingprediction-s based on time series data, since there can be lags of unknown duration between important events in a time serieswhich are very well designed to deal with the vanishing gradient problem that can be encountered when training traditiona Rnn's .

Comparision Of Algorithms

Test case ID	Test case name	Test case Desc	Test steps			Test case status	Test priority
			step	expected	actual		
01	Upload dataset	Verify either dataset loaded or not	If dataset is not uploaded	We cannot perform further operations on the dataset	Uploaded data set	high	high
02	Support vector regression	Verify either Support vector or not	If not process	Run successfully or not	Run successfully	low	high
03	Random forest regression	Verify either support or not	Run random forest	Run successfully or not	Run successfully	low	high
04	Gradient boosting regression	Verify either support or not	Run gradient boosting	Run successfully or not	Run successfully	low	high
05	LSTM Deep learning	Verify either support or not	Run LSTM alg	Run successfully or not	Run successfully	high	high
06	Upload test values	Verify either data test values loaded or not	If test values is not there	We cannot perform operations on the test	Uploaded test values	high	high

RESULTS

DATE,TIME, sensor id,HYBX,HYBY,HYBZ,HYBF
2020-01-14,00:00:00.000,014,46.13,4.16,71.96,43615.62
2020-01-14,00:01:00.000,014,46.25,4.13,71.95,43615.72
2020-01-14,00:02:00.000,014,46.30,4.07,71.95,43615.78
2020-01-14,00:03:00.000,014,46.38,3.99,71.89,43615.81

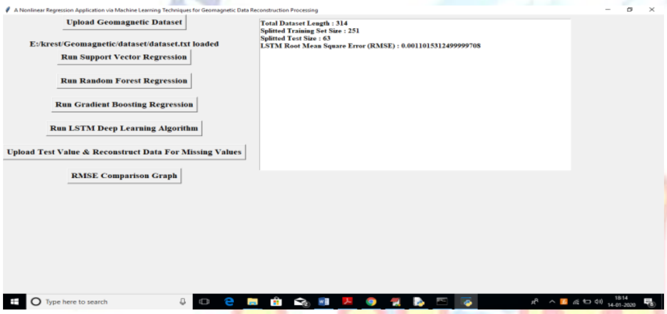
Above dataset obtained from HYDERABAD Area Sensor so its ID contains HYB and HYBX is the latitude and HYBY is the longitude and others values are the earth magnetic data. Along with this data we can see sense date and time with sensor id. In above dataset sensor is configure to sense value every 1 minute and if we want to have value in half minute then that value is missing. For example in above dataset

First record time = 00:00:00 and it target value = 43615.62

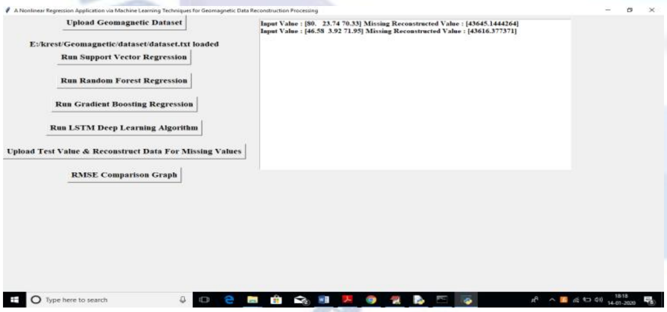
Second record time = 00:01:00and it target value = 43615.72

missing value at time = 00:00:30 we want to have 00:00:30 and this missing value we can obtained

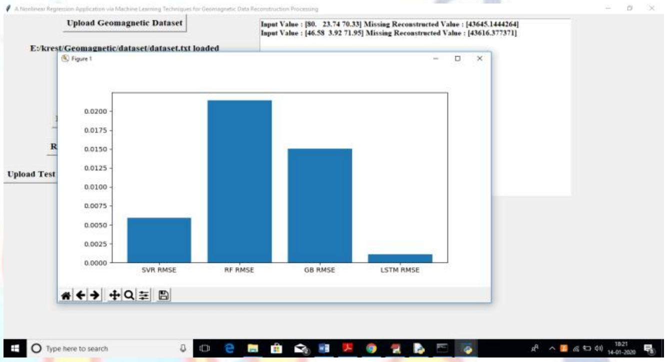
by applying regression algorithms



In above screen we can see LSTM RMSE error rate, now click on 'Upload Test Value & Reconstruct Data For Missing Values' button and upload 'test_dataset' file and this file contains some values whose target value is missing and this application will predict target value for those missing values. See below records from test file.



.In above screen we got missing fourth value which we called as reconstructed or predicted value. Similarly u can add more intervals values in test dataset file and get its missing target value. Now click on 'RMSE Comparison Graph' button to get below graph.



DISCUSSION

we proposed both the machine learning and deep learning methods for geomagnetic data reconstruction. A hidden relationship (continuous hyperplane) can be determined using sufficient exemplified training sets, from which the missing data can also be derived. We present the RNN method to improve the performance of the classic machine learning method (SVM, gradient boosting, and random forests). Besides, the LSTM-based approach allowed us to avoid previous drawbacks in the existing reconstruction methods, and it is generally applicable to different data sets. 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 higher than the traditional method. The deep learning methods showed the high accuracy, while it still can be improved as they are being further developed and many interesting ideas are emerging. In the future work, we will also investigate the application of ensemble methods for geomagnetic data processing, which can integrate multiple machine learning models and assign each model with a weighting factor, such that different methods are assigned to apply on their better performed data, so the accuracy can be improved.

References

[1] S. Khomutov, V. Sapunov, A. Denisov, D. Savelyev, and I. Babakhanov, "Overhauser vector magnetometer POS-4: Results of continuous measurements during 2015–2016 at geophysical observatory 'Paratunka' of IKIR FEB RAS, Kamchatka, Russia," in Proc. E3S Web Conf., vol. 11.Les Ulis, France: EDP Sciences, 2016, Art. no. 00007.

[2] H. Dong, H. Liu, J. Ge, Z. Yuan, and Z. Zhao, "A high-precision frequency measurement algorithm for FID signal of proton magnetometer,"IEEE Trans. Instrum. Meas., vol. 65, no. 4, pp. 898–904, Apr. 2016.

[3] R. M. Winslow et al., "Observations of Mercury's northern cusp region with MESSENGER's Magnetometer," Geophys. Res. Lett., vol. 39, no. 8, p. L08112, 2012.

[4] F. Lhuillier, J. Aubert, and G. Hulot, "Earth's dynamo limit of predictability controlled by magnetic dissipation," Geophys. J. Int., vol. 186,no. 2, pp. 492–508, 2011

[5] G. Hulot, F. Lhuillier, and J. Aubert, "Earth's dynamo limit of predictability," Geophys. Res. Lett., vol. 37, no. 6, p. L06305, 2010.

[6] M. Dumberry and C. C. Finlay, "Eastward and westward drift of the Earth's magnetic field for the last three millennia," Earth Planet. Sci. Lett., vol. 254, nos. 1–2, pp. 146–157, 2007.

[7] M. Korte, F. Donadini, and C. Constable, "Geomagnetic field for 0–3 ka: 2. A new series of time-varying global models," Geochem., Geophys.,Geosyst., vol. 10, no. 6, p. Q06008, 2009.

[8] I. Wardinski and M. Korte, "The evolution of the core-surface flow over the last seven thousands years," J. Geophys. Res., Solid Earth, vol. 113,no. B5, p. B05101, 2008.

[9] M. Korte and R. Holme, "On the persistence of geomagnetic flux lobes in global Holocene field models," Phys. Earth Planet. Interiors, vol. 182, nos. 3–4, pp. 179–186, 2010.

[10] M. Korte, C. Constable, F. Donadini, and R. Holme, "Reconstructing the Holocene geomagnetic field," Earth Planet. Sci. Lett., vol. 312, no. 3, pp. 497–505, 2011.

[11] A. Nilsson, R. Holme, M. Korte, N. Suttie, and M. Hill, "Reconstructing Holocene geomagnetic field variation: New methods, models and implications," Geophys. J. Int., vol. 198, no. 1, pp. 229–248, 2014.

[12] L. Kapper, F. Donadini, V. Serneels, E. Tema, A. Goguitchaichvili, and J. J. Morales, "Reconstructing the geomagnetic field in West Africa: First absolute intensity results from Burkina Faso," Sci. Rep., vol. 7, Mar. 2017, Art. no. 45225.