GPR data processing methods based on machine learning algorithm to detect the backfill grouting of shield tunnel

Li Zeng[a,b], Xiongyao Xie[b], Biao Zhou[b]*
[a]Université catholique de Louvain, [b]Tongji University

Summary

Shield tunnel method is currently the most important method for tunnel excavation in soft soil areas. With the construction and operation of a large number of tunnels in China, a certain degree of structural disease have appeared in the tunnels. How to control the settlements of tunnels and take corresponding measures to ensure the safety of operations has attracted great attention from design and operation management departments. In this chapter, FDTD numerical simulation of backfill grouting by GPR is performed. The GPR images of grouting layers with different thicknesses of 400MHz antenna and 900MHz antenna were simulated respectively, and the theory of integrated learning XGBoost is established. The XGBoost model was trained with numerical modeling data to obtain classification models of grouting layers of different thicknesses under 400 MHz and 900 MHz GPR, and the results showed that the GPR data could be classified and pattern recognized well under this circumstance.

Introduction

The detection of backfill grouting of shield tunnel structures according to surface-based geophysical reflection data is an interesting question in ground penetrating radar (GPR) research. Many studies have carried out to solve the feasibility of this problem effectively. However, since the radar images recorded from the reflected echoes are not explicit geometric shapes of the underground structures, the geophysical data collected from either numerical or field studies are seriously subjected to geological conditions and electromagnetic interference of varying degrees. According to the literatures found, the resolution and accuracy of the results are still inadequate to meet the requirements of engineer detection. With the continuous development and innovation of machine-learning methods, there are more and more cases to process geophysical reflection data applying machine-learning methods[1]. These precedent cases[2] showed us the possibility of improving the detection accuracy of backfill grouting of shield tunnel lining based on GPR.

The focus of the backfill grouting construction is whether the grouting is plump and adequate. According to the field engineering experience of the Shanghai North Cross Tunnel, the quality of backfill grouting can be evaluated and classified according to certain thickness of the grout. From the perspective of pattern recognition and machine learning, this is a three-category-classification problem, that is, the A-scan corresponding to under-compensated grouting, normal grouting and super-compensated grouting are possible to be separated.

Theory and/or Method

Ensemble learning[3] is a cutting-edge algorithm in the field of algorithm competition and machine learning. Almost all applications of machine learning related algorithms can be used. Ensemble learning is not a single machine learning algorithm, but a combination of multiple machine learners to complete data mining through certain algorithm rules. For the training set data, we apply several individual learners to acquire a certain result, set up a combination rules to combine, and finally form a strong learner.

Gradient boosting decision tree (GBDT)[4][5] is an important member of the Boosting family of ensemble learning. Its weak learners are limited to CART regression tree models. The GBDT iterative method also uses a forward distribution algorithm.

As an efficient and engineered implementation of GBDT, XGBoost [6] is a novel statistic-learning algorithm with an extremely high limitation, therefore it is more popular in algorithm competitions. In short, compared with the original algorithm GBDT, XGBoost is optimized from the following three aspects: Optimization of the algorithm itself; Optimization of algorithm operation efficiency; Optimization of algorithm robustness.

Algorithm evaluation is the basis for achieving reasonable selection of parameters that support the XGBoost algorithm. Therefore, the concept of cross validation (CV)[7]can be introduced during the training of the XGBoost algorithm. Cross validation is one of the most commonly used technique for estimating the generalization error of regressions or classifier.

In this paper, FDTD numerical simulation was applied to analyze the backfill grouting of the Shanghai North Cross shield tunnel by GPR. The ground penetrating radar images of grouting with different thicknesses of 400MHz antenna and 900MHz antenna were simulated respectively. The XGBoost model was trained with FDTD modeling data to obtain classification models of grouting layers of different thicknesses at 400MHz and 900 MHz GPR. During the training process, cross validation and parameters' grid search were also involved in the selection of proper parameters for XGBoost models.

Besides XGBoost, which based on decision trees and ensemble learning algorithms, we also applied traditional Multilayer Perceptron (BP neural network) with six layers neural network as a compared algorithm trained by same modeling data.

Examples (Optional)

In this chapter, GprMax 3.0[8] software would be used to perform forward modeling of the backfill grouting of shield tunnel.

According to the backfill grouting detection characteristics, a two-dimensional FDTD forward modeling block diagram and PML absorption boundaries [9] were established. The 8 layers on the left, right, and bottom are set as PML absorbing layers. This program was compiled in python environment.

In the forward model, the size of the discrete grid plane is 400*400; the maximum vertices of the boundaries were (400, 400) and the minimum vertices were (0, 0); Suppose the incident wave was perpendicular to the interface (θ = 0 °) in the forward numerical model. The dielectric constant of the concrete linings was 6.5 and the dielectric constant of the grouting slurry was 36. It was used to simulate the initial grouting and the dielectric of the sand body. The constant is

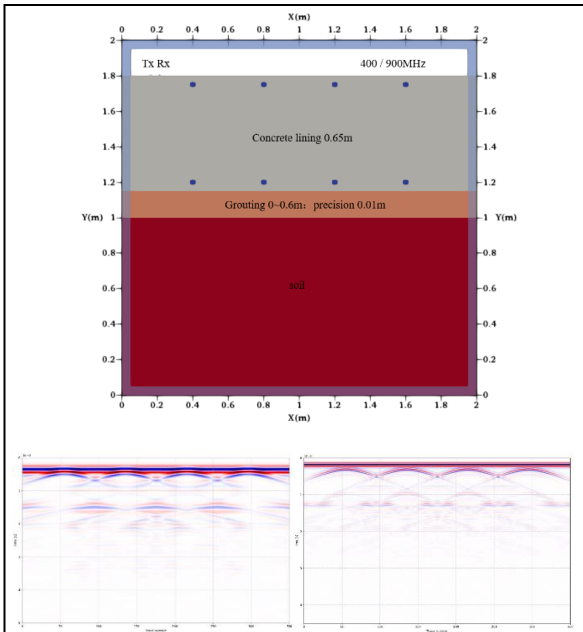

Figure 1: GPRMAX3.0 grouting models with 400MHz&900MHz antennas

20, and the dielectric constant of air is set at 1. Since the polarization direction of the antenna was the vertical tube downward, the incident wave could be set to be perpendicular to the interface (θ = 0 °). In this paper, two different frequency GPR, 400MHz and 900MHz, were tested in the numerical experiments. In order to compare with future model experiments, the time window selected for the 400MHz radar is 50ns, and the 900MHz radar 45ns. The distance between the two kinds of radar's transmitting and receiving antennas is 50mm, and the track spacing was 5mm.The partial results of the numerical experiments were shown in Figure 1.

The purpose of GPR data pre-processing is to suppress noises, enhance the signals, and improve the signal-to-noise ratio. Compared with the measured data, the forward simulation data has a higher signal-to-noise ratio, but there is still a certain level of interference from direct waves, surface reflections, and multiples. Therefore, some pre-processing should be executed before the identification to improve the data quality.

(1) Go drifting. Make the probability distribution of the amplitude of the signal symmetrical.
(2) Butterworth band-pass filtering. Remove the interference signals of biased frequency in the data.
(3) Moving average denoising. Moving average is a common denoising method for signal processing. In this chapter, n = 3 is selected for filtering based on experience. The above methods can be implemented by radar signal processing software named Reflexw.
(4) Normalized. [-1, + 1] normalization method is used to linearly normalize the data in the matrix to the range of [-1, + 1].

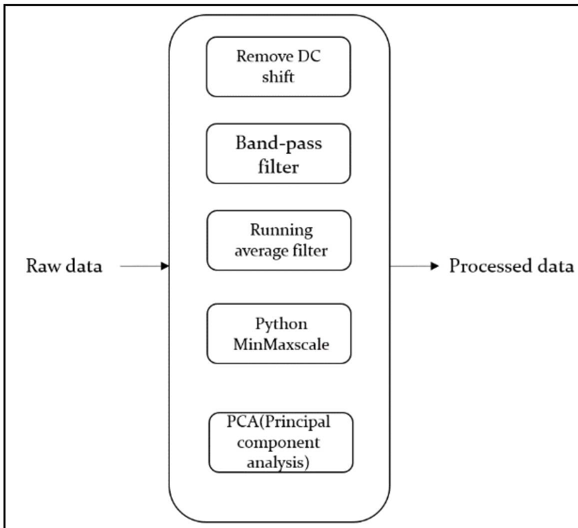

Figure 2: Preprocessing procession

(5) PCA(Principal component analysis). The sampling point of the forward analog signal is 1024, that is, an A-scan signal consists of 1024 electromagnetic time-domain signal points. However, for the XGBoost, the dimensions of the signal should be as small as possible. PCA belongs to unsupervised-machine-learning algorithms, which has been widely used in data compression, elimination of data redundancy and other scopes.

The process of pre-processing before the involvements of machine learning(supervised) is shown in Figure 2.

Radar images cannot be used as input data and samples for machine learning directly. Before image recognition, feature extraction must be performed on the images to extract the signal features of the super-compensated grouting, normal grouting and under-compensated grouting. Taking all kinds of feature extraction methods into consideration, the time-domain feature extraction of the GPR signals is relatively easy, and the amplitude, shape, and position of the signal itself can also represent some properties of the measured targets. Therefore, this paper extracts signals from GPR time domain images as features.

In this chapter, XGBoost recognition program used Scikit-Learn and XGBoost modules based on python locale. The general instructions to use these modules were: firstly constructed a training set in the format required by the python modules; then trained XGBoost in the training set to obtain the XGBoost model; the trained XGBoost model could be used to identify samples outside the training set.

After performing numerical experiment modeling for 400MHz and 900MHz antennas, training samples were collected from backfill grouting training data of different antenna frequencies. 5600 super-compensated grouting training samples, 7000 normal grouting training samples and 5600 under-compensated grouting training samples were obtained. Because the number of samples in the three training sets is approximately even, you can directly select the pre-processed signals and then generate sample labels corresponding to the training sample matrix, with matrices of sizes 5600*1024, 7000*1024, and 5600*1024. In the python program, the ratio of the training set, test set, and verification set is 8: 1: 1 and the built-in function can randomly separate these three samples. In this way, a training set of backfill grouting for XGBoost is obtained.

The problem based on the grouting thickness could be regarded as classification from the perspective of machine learning. For the GPR data of different grouting thicknesses, the accuracy of grouting thickness under XGBoost algorithm should be explored by integration learning. In this part of the data set, samples with 0.01m grouting thickness precision sampling were selected for training and verification, meaning that the accuracy of recognition is 0.01m.

By analyzing the accuracy obtained by XGBoost training, we can draw the following conclusions:

1. On a data set with a thickness accuracy of 0.01m backfill grouting, the classification training set and verification set achieved an accuracy rate of over 96%, which basically met the requirements in engineering fields.

2. The GPR data collected by the 400MHz and 900MHz antennas both perform well in XGBoost pattern recognition. The accuracy of the 900MHz forward wall grouting accuracy is 0.01m. The accuracy of the forward simulation data is slightly lower than that of 400MHz, but it still met the accuracy requirement. Due to the 900MHz GPR's high resolution but low penetration, it was generally considered difficult to distinguish between the backfill grouting thickness recognition of the Shanghai North Crossing Tunnel. However, due to the powerful performance of

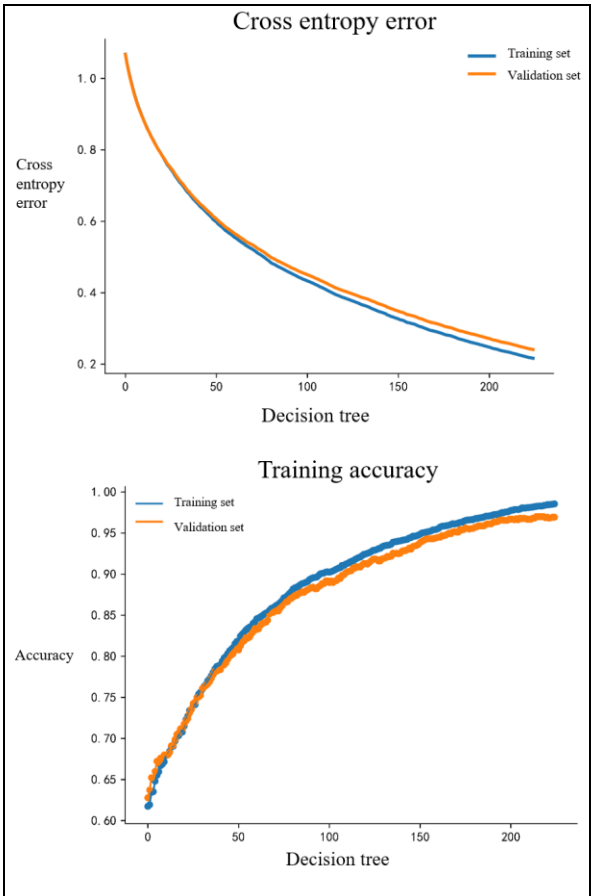

Figure 4: XGBoost model training error and accuracy

machine learning, the models trained on the dataset perform better.

3. On the 400MHz & 900MHz GPR simulation data set, the data set training showed that the BP neural network was weak on the data, with the training set reaching 53.7%, and the verification set 53.0%. The three types of data could not be trained based on multilayer perceptron. The overall training drove into gradient disappearance. Compared with the previous data using dimensionality reduction, the training time is too long, one training takes about 15-20 minutes, which is a lot slower than XGBoost (3-5min). Therefore, in this case, XGBoost is a better method on processing GPR data collected from backfill grouting.

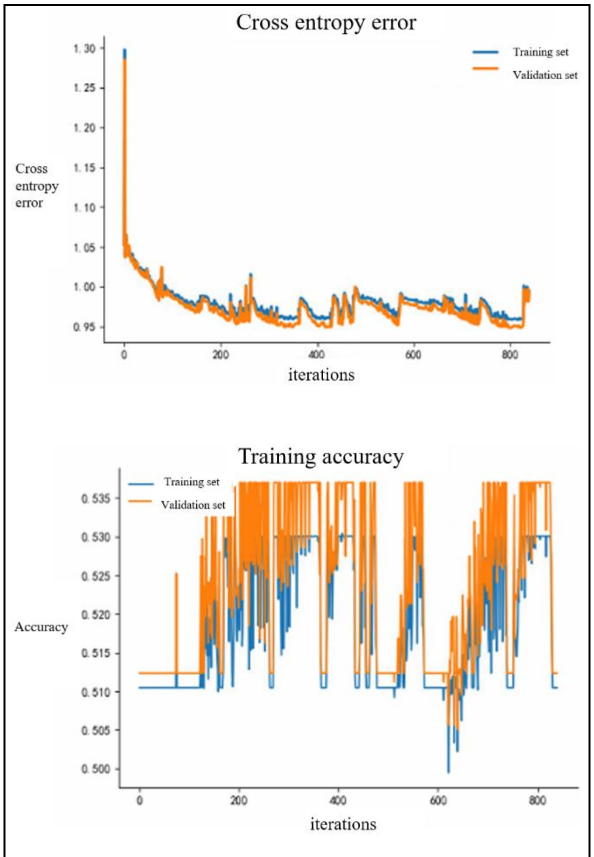

Figure 5: Multilayer Perceptron model training error and accuracy

Conclusions

In this chapter, FDTD numerical simulation of backfill grouting of shield tunnel by GPR was performed. The ground penetrating radar images of grouting layers with different thicknesses of 400MHz antenna and 900MHz antenna were modelled respectively. The XGBoost model was trained with modeling data to obtain classification models of backfill grouting of different thicknesses at 400 MHz and 900 MHz GPR. The main conclusions are as follows:

1. On a data set with a thickness accuracy of 0.01m backfill grouting, the classification training set and verification set achieved an accuracy rate of over 96%, which basically met the requirements in engineering fields.

2. Compared with traditional multilayer perceptron, in this case, XGBoost was a better method on processing GPR data collected from backfill grouting.

3. XGBoost model also met the requirements of training speed, with the results generated within 3-5 minutes. The engineers can get the real-time classification results in the construction sites if the XGBoost model is integrated on the system of GPR data processing modules.

References

[1] Xie X Y，Li P，Qin H, et al. GPR identification of voids inside concrete based on the support vector machine algorithm [J]. Journal of Geophysics and Engineering, 2013, 10(3): 034002

[2] Lei W, Hou F, Xi J, et al. Automatic hyperbola detection and fitting in GPR B-scan image [J]. Automation in Construction, 2019, 106, 102839.

[3] Romov P A. RecSys Challenge 2015: ensemble learning with categorical features[C]// 2015.

[4] Mitchell T M. Key ideas in machine learning[G]. Machine Learning. New York，U.S.: McGraw Hill，1997, 1-11.

[5] Friedman J H. Greedy function approximation: A gradient boosting machine [J]. Annals of Statistics, 2001, 29(5):1189-1232.

[6] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System[C]. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016, 785-794

[7] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]. 14th International Joint Conference on Artificial Intelligence, San Mateo, Canada, 1995: 1137~1143

[8] C. Warren, A. Giannopoulos, I. Giannakis. gprMax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar[J]. Computer Physics Communication, 2016, 209, 163–170.

[9] Berenger J P. Making use of the PML absorbing boundary condition in coupling and scattering FDTD computer codes [J]. IEEE Transactions on Electromagnetic Compatibility, 2003, 45(2):189-197.

