A discrete element modeling of rock and soil material based on the machine learning

Bing Yuan\textsuperscript{1}, Chun Liu\textsuperscript{1,2}, Yan Qin\textsuperscript{1}, Tiansheng Zhang\textsuperscript{1} and Xiaodong Ma\textsuperscript{2}

\textsuperscript{1}School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China.
\textsuperscript{2}Department of Earth Sciences, ETH Zürich, 8092, Switzerland

\textcopyright Chun Liu

chunliu@nju.edu.cn

Abstract. The discrete element modeling of rock and soil mass usually relies on tedious parameter adjustment and mechanical property testing operations to obtain appropriate mechanical parameters of the element. In this paper, based on the linear elastic contact theory, a discrete element model of rock samples was established based on discrete element software MatDEM with high-performance numerical analysis, and the actual mechanical properties of the model samples were predicted using a machine learning method, XGBoost algorithm. A certain amount of sample data sets was generated, and through numerical experiments, training, validation and test operation of the rock mechanical properties (compressive strength, Young's modulus, Poisson's ratio and tensile strength), the numerical model quantitative relationship between the actual and input mechanical properties was explored, showing the following results: (1) Based on the characteristics of the model 12000 rock samples and correlation analysis, the tag data shows that the input mechanical properties are positively correlated with the actual mechanical properties; (2) With the increase of the number of samples in the sample training data set, the training average error becomes smaller and smaller, and the error distribution becomes more and more stable. (3) The training results converge to the set threshold, and the prediction errors of mechanical parameters are all less than 4%, and the error distribution is stable, which proves that the machine learning method can quickly and accurately establish the rock model with specific mechanical properties.

1. Introduction

The rock-soil mass is relatively continuous on the macroscopic level, while on the microscopic level it is a systematic structure composed of a series of particles, pores and cracks. The microscopic discrete and discontinuous problems are difficult to solve by conventional methods based on continuum mechanics. Discrete element method builds models by accumulating and cementing particles, which can naturally simulate the discontinuity and inhomogeneity of rock and soil mass, and provides an effective method for exploring the micromechanism of rock and soil mass. In recent years, with the improvement of computing power, the discrete element method has gradually been applied to the study of large-scale problems such as geological disasters, geotechnical engineering, rock mechanics and structural evolution.
The mechanical properties of the discrete element accumulation model are affected by multiple factors such as particle properties, stacking process and cementation, and it is usually impossible to directly obtain a discrete element model with specified mechanical properties. Discrete element method modeling often relies on a series of adjustment and testing operations to obtain the accumulation model of the specified mechanical properties, which is commonly referred to as the "trial-and-error method[1]". This method makes the macro-mechanical response of the discrete element model reach the ideal state or consistent with the laboratory test results by continuously adjusting the micro-element mechanical parameters. For example, Heekwang Lee and Seokwon Jeon used the trial-and-error method to calibrate the mechanical parameters of the granite numerical model, and numerically simulated the coalescence characteristics of granite[1].

Although the trial-and-error method is universal, its parameter adjustment process is complicated, the threshold is high and the time is long, which greatly limits the application of discrete element method in engineering. Therefore, based on the simulation or test results, some scholars statistically analyzed the relationship between the macroscopic mechanical properties and the microscopic parameters, summarized the empirical formula for the conversion between them, and applied it to the discrete element rapid modelling[2]. For example, Liu et al. deduced the conversion formula between parameters of discrete element 3D dense accumulation model and rock mechanical properties, and took the quartzite model as an example to verify the correctness of the conversion formula, which provided important guidance for discrete element quantitative modeling[3]. These methods show good modeling accuracy and efficiency.

In recent years, with the further development of artificial intelligence, machine learning algorithm has been gradually applied to the research of geotechnical engineering. Machine learning method has strong data mining ability, and it has a very good application prospect in discrete element macro and micro modeling, especially in automatic and efficient prediction and modeling.

This article introduces the machine learning algorithm XGBoost (eXtreme Gradient Boosting). The numerical model of rock and soil mass was established based on the linear elastic contact theory with the high-performance discrete element software MatDEM independently developed by Nanjing University. A series of numerical analysis experiments, model training, testing and validation operations were carried out. With the powerful data mining ability of machine learning algorithm, the quantitative relationship between the actual mechanical properties of rock samples and the input mechanical properties of numerical models is explored.

2. Methods and principles

2.1. Linear elastic contact model

In this paper, a discrete element linear elastic contact model is adopted, as shown in Figure 1. It is assumed that the elements rely on springs to contact each other and produce force action. The normal force \(F_n\) and normal deformation \(X_n\) between the two elements can be simulated by the normal spring, as shown in Figure 1a.

![Figure 1. Linear elastic contact model (a) The normal contact model; (b) The tangential contact model](image)
In equation (1), \( K_n \) is the normal stiffness; \( X_n \) is the normal relative displacement; \( X_b \) is the fracture displacement. When the normal relative displacement between the two units exceeds its breaking displacement, the spring between the units will break and the spring tension will disappear (Equation 1c). When the two return to the compressed state, there is still pressure between the units (Equation 1b). In the same way, the shear force \( (F_s) \) and shear deformation \( (X_s) \) between the two elements can be simulated by a tangential spring, as shown in Figure 1b.

\[
F_n = \begin{cases} 
K_nX_n, & X_n < X_b \text{ intact bond (a)} \\
K_nX_n, & X_n > 0 \text{ broken bond (b)} \\
0, & X_n = 0 \text{ broken bond (c)} 
\end{cases}
\] (1)

In equation (2), \( K_s \) is the shear stiffness; \( X_s \) is the shear displacement (Figure 1b). The failure criterion of the spring in the tangential direction is based on the Mohr-Coulomb criterion:

\[
F_s = K_sX_s
\] (2)

In equation (3), \( F_{s\text{ max}} \) is the maximum shear force; \( F_{s0} \) is the shear resistance between elements; \( \mu_p \) is the friction coefficient between elements. In the Mohr-Coulomb criterion, the maximum shear force between elements is related to the initial shear resistance. The shear force is positively correlated with the normal force \( F_n \), which increases with the increase of the normal force. When the tangential force exceeds the maximum shear force, the connection of the tangential spring will break and the two units will begin to slide relatively to each other. The sliding friction force between the units is \( -\mu_p F_n \).

2.2. XGBoost algorithm

XGBoost is an efficient and widely used promotion tree extensible machine learning method. It has the advantages of preventing overfitting, reducing computation, parallel optimization and customizable loss function, etc., and has a wide range of applications in drilling engineering, geotechnical engineering, material engineering, pollution control and medicine[4-6].

2.2.1. Objective function.

Assume that \( X \) and \( Y \) are input parameters and output parameters respectively, and \( Y \) is a continuous variable. Given a data set:

\[
D_n = \{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}
\] (4)

The regularization term is introduced to optimize the experience loss and complexity of the objective function to reduce overfitting. The complexity is defined by the following formula:

\[
\Omega(f) = \eta T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2
\] (5)

In equation (5), \( \omega \) is the output fraction of each node, \( T \) represents the number of leaf nodes, \( \eta \) and \( \lambda \) represent the regularization coefficient. According to the forward distribution algorithm (assuming \( T \) is the number of trees), the structure of the first t-1 tree is a constant, and the constant term has no influence on the optimal value point of the objective function. The mean square error is used as the loss function, then the function is expressed as:

\[
Obj(f) = \sum_{t=1}^{n} \left[ 2 \left( \hat{y}_i^{(t-1)} - y_i \right) f_t(x_i) + f_t(x_i)^2 \right] + \Omega(f_t) + C
\] (6)

The Taylor expansion formula is used to simplify the function and make subsequent calculations convenient. Define the first-order derivative \( g_t \) and the second-order derivative \( h_t \):

\[
g_t = \frac{\partial}{\partial y_i^{(t-1)}} (\hat{y}_i^{(t-1)}), h_t = \frac{\partial^2}{\partial y_i^{(t-1)^2}} (\hat{y}_i^{(t-1)})
\] (7)

Bring equation (7) into equation (6) for integration, define \( I_t \) as the combination of all samples on the j-th leaf node, and then bring in according to the penalty term of equation (6), \( G_j = \sum_{i \in I_t} g_t, H_j = \sum_{i \in I_t} h_t \). Derivation to obtain the final objective function:
\[ \text{Obj} = -\frac{1}{2} \sum_{j=1}^{T} \frac{c_j^2}{H_j + \lambda} + \eta T \] (8)

3. Discrete element modeling and numerical experiments

3.1. Discrete element modeling of sample

![Figure 2](image)

Figure 2. Training model (a) The geometric model; (b) Model particle size distribution

Discrete element modeling and numerical simulation are operated by high-performance discrete element software MatDEM. It has been successfully used in landslide investigation, clay cracking, land subsidence and other studies\(^{[7-8]}\).

The sandstone cuboid model was established through the gravity accumulation operation of MatDEM software, as shown in Figure 2(a). The size of the model is: length 50mm, width 50mm, and height 120mm. The model includes three parts: sample, lower pressure plate and upper pressure plate. Pressure can be generated by applying physical force to the upper pressure plate. The sample part has a total of 8822 units. The specific particle size distribution is shown in Figure 2(b).

3.2. Material assignment and testing

After the accumulation model was established, the macro-mechanical properties of sandstone were given to the model. The actual macroscopic mechanical properties of the model set in this study include Young’s modulus \(E\), Poisson’s ratio \(\nu\), tensile strength \(T_u\), compressive strength \(C_o\), internal friction coefficient \(\mu_i\) and density \(\rho\). The macroscopic input mechanical properties of rocks are obtained by conventional rock mechanical properties test methods. The macro-micromechanical parameter conversion formula proposed by Liu et al.\(^{[3]}\) is used to determine the micromechanical parameters between elements. The conversion formula of macro and micro mechanical parameters is based on the tightly packed discrete element model. For the random packed model, due to the reduction of the coordination number, its mechanical properties are usually reduced to 10%-40% of the input value. For this reason, the default mechanical properties of MatDEM software are set to expand the scale coefficient. Young's modulus, Poisson's ratio, compressive strength and tensile strength are multiplied by the coefficients 2.7, 0.8, 6 and 6.5 respectively, and then substituted into the transformation formula to obtain the micro-mechanical parameters of the element, which are automatically assigned to each element.

After completing the material settings, numerical experiments were performed to obtain the actual mechanical properties of the model: uniaxial compression simulation tests to test the Young’s modulus, Poisson’s ratio and compressive strength of the discrete element model; uniaxial tensile tests to test the tensile strength of the discrete element model; density is obtained by dividing the mass of the tightly packed model by the volume.
11th Conference of Asian Rock Mechanics Society

IOP Conf. Series: Earth and Environmental Science 861 (2021) 032015
doi:10.1088/1755-1315/861/3/032015

3.3. Correlation analysis of sample data sets

12,000 model samples with the same size and particle size distribution were established and assigned different macroscopic mechanical properties. Young's modulus is distributed between 0.1 - 402 GPa. Poisson's ratio ranges from 0.01 - 0.2. The tensile strength is distributed between 0.11-206 MPa. The compressive strength is distributed between 27 - 1456 MPa. The internal friction coefficient is distributed between 0.3-2. The density distribution is between 1750 - 2934 kg/m³. The actual mechanical properties corresponding to 12,000 samples were obtained through numerical tests, and the training data set composed of 12,000 input mechanical parameters and test mechanical parameters samples was obtained.

A preliminary correlation analysis was conducted on the input mechanical parameters and the actual mechanical parameters before model training. Pearson correlation coefficient was adopted to represent the degree of linear correlation between the two variables, and its calculation formula was as follows:

\[ r(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \]  \hspace{1cm} (9)

Figure 3. Correlation coefficient between input mechanical properties and actual mechanical properties: (a) Young's modulus; (b) Poisson's ratio; (c) Tensile strength; (d) Compressive strength

In equation (9), \( X \) and \( Y \) represent two variables, \( \text{Cov} \) is the covariance, and \( \text{Var} \) is the variance. The value of \( r \) is \([-1,1]\), less than 0 indicates negative correlation, greater than 0 indicates positive correlation, and the closer to 0, the smaller the correlation. According to formula (9), the correlation coefficient between the input mechanical properties and the actual mechanical properties is calculated, and a scatter diagram showing the correlation is drawn. As shown in Figure 3.
The correlation coefficients between actual mechanical property and input Young's modulus, Poisson's ratio, tensile strength and compressive strength are 0.76, 0.9, 0.91 and 0.94, respectively, showing a good positive correlation. For the same input parameter value (e.g., Young's modulus is equal to 40GPa), the actual mechanical parameters of the model established are not unique (Figure 3). Figure 3b is a scatter diagram of the input Poisson's ratio and the actual Poisson's ratio. The maximum Poisson's ratio of the transformation formula is 0.2, but the actual Poisson's ratio of the model varies between 0.15 and 0.30 due to the influence of other mechanical properties.

In summary, it is possible to obtain materials similar to the input mechanical properties by using the conversion formula and the mechanical property expansion coefficient. However, due to the complexity of the discrete element method modeling, the mechanical parameters of the model are affected by many factors, and it is difficult to accurately obtain the quantitative relationship between the input mechanical properties and the actual ones, so machine learning algorithms are introduced to train and establish the quantitative relationship between the two.

4. XGBoost model training and result analysis

4.1. Model training

The sample data matrix was loaded, containing feature and label data. The four labels of actual Young's modulus, Poisson's ratio, tensile strength and compressive strength correspond to the four characteristics of input Young's modulus, Poisson's ratio, tensile strength and compressive strength respectively. One sample for each behavior corresponds to one feature for each column of training data. In order to avoid the difference of orders of magnitude between different features affecting the training accuracy and making the results difficult to converge, 0~1 normalization is required for features and labels:

\[ x = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (10) \]

In equation (10), \( x \) is the current data, \( x_{\text{min}} \) is the minimum value in the data, \( x_{\text{max}} \) is the maximum value in the data, and \( x \) is the normalized data. The 12,000 sample data were divided according to rows, among which 1200 (10%) were randomly selected as the test set, 1200 (10%) were randomly selected as the verification set, and the remaining 9,600 (80%) as the training set.

The XGBoost algorithm model was loaded, and the processed sample data set was input into the XGBoost algorithm for training. The root mean square loss function is used for evaluation during training. When the validation set error does not decrease for 10 consecutive trainings, the training ends. After the training, the data was denormalized and restored, and then the average relative error was used to calculate the error of the predicted value of the test set. The formula for calculating the error of a single sample was as follows:

\[ \text{error} = \frac{|Y_{\text{pre}} - Y_{\text{test}}|}{Y_{\text{test}}} \times 100\% \quad (11) \]

In equation (11), \( Y_{\text{pre}} \) is the predicted value and \( Y_{\text{test}} \) is the actual value. Error convergence threshold value was set as 4% to judge whether the training error converges to the threshold. If the error does not converge, the algorithm parameters \( n_{\text{estimator}} \), \( \text{max depth} \) and \( \text{learning rate} \) should be re-selected and the training should be carried out again until the error converges. Through continuous trial calculation and verification, the actual compressive strength, tensile strength, Poisson's ratio and Young's modulus were obtained with the minimum error under the parameter settings shown in Table 1, and then the trained XGBoost model was obtained.

**Table 1.** Optimal parameters for compressive strength, Young's modulus, Poisson's ratio and tensile strength

|                      | max depth | learning rate | n estimators |
|----------------------|-----------|---------------|--------------|
| Young's modulus      | 11        | 0.1           | 150          |
| Poisson's ratio      | 12        | 0.11          | 110          |
| Tensile strength     | 13        | 0.1           | 110          |
| Compressive strength | 13        | 0.11          | 100          |
4.2. Model training results
After the model was trained, correlation analysis was performed on the predicted results of the actual compressive strength, tensile strength, Poisson's ratio and Young's modulus of the test set prediction results, shown in Figure 4. The errors of predicted results are presented in Table 2 where the average error indicates the overall accuracy, and the error standard deviation refers to its distribution. Compared with the correlation coefficient between the input mechanical properties and the actual mechanical properties in Figure 3, it is found that the predicted values in Figure 4 are all distributed along \( Y = X \). After training, the correlation coefficient \( r \) between the predicted values and the actual values is 0.99 for all four groups.

![Prediction results](image)

**Figure 4.** Prediction results: (a) Young's modulus; (b) Poisson's ratio; (c) Tensile strength; (d) Compressive strength

|                     | Young's modulus | Poisson’s ratio | Tensile strength | Compressive strength |
|---------------------|-----------------|-----------------|------------------|----------------------|
| Average error       | 2.9%            | 1.6%            | 3.8%             | 1.5%                 |
| Error standard deviation | 7.8%          | 4.3%            | 9.4%             | 4.8%                 |

5. Conclusions
Based on the powerful data mining capability of machine learning XGBoost algorithm, combined with the high-performance discrete element software MatDEM, the discrete element model of rock is
established by using the linear elastic contact theory. A training model to train the input mechanical properties of rock model samples was built, and the quantitative relationship was explored between the input mechanical properties of rock and the actual ones. Through numerical experiments, training, verification and testing operations, the following conclusions are drawn:

1) Using XGBoost algorithm for training 12000 rock samples of discrete element tag data, the quantitative relationship was established between physical and actual mechanical properties, 12000 features of rock mechanics properties of the sample data were also generated. Correlation analysis was carried out on the features and labels data, and there are highly positive correlations of the four labels of actual Young’s modulus, Poisson’s ratio, tensile strength and compressive strength with the input Young’s modulus, Poisson’s ratio, tensile strength and compressive strength.

2) The training results of 12,000 data sets are taken as the final training results, and the prediction errors of the actual Young's modulus, Poisson's ratio, tensile strength and compressive strength are all less than 4%, which can be predicted that the errors will be smaller when the sample size is larger.

3) All the training results converge to the set threshold, and the average error of all the predictions is less than 4%. The average error of Poisson's ratio and compressive strength can reach less than 2%, and the error distribution is stable. In the case of ensuring the modeling accuracy, the discrete element modeling efficiency is greatly improved, and tedious parameter adjustment process is avoided. It is proved that the XGBoost algorithm is effective in discrete element modeling, and the machine learning method can be used to quickly and accurately establish rock models with specific mechanical properties.

References

[1] Lee H and Jeon S. International Journal of Solids and Structures 48.6(2010): 979-99. doi:10.1016/j.ijsolstr.2010.12.001.

[2] Duan, K.*, Kwok, F.† and Ma, X. DEM simulations of sandstone under true triaxial compressive tests, Acta Geotechnica 12 (3): 495-510(2017), doi.org:10.1007/s11440-016-0480-6

[3] Liu C, Xu Q, Shi B, Deng S and Zhu H H. Computers and Geosciences 103.(2017): 12-20. doi:10.1016/j.cageo.2017.03.003.

[4] Zhong R Z, Johnson R and Chen Z W. International Journal of Coal Geology . 220.(2020):. doi:10.1016/j.coal.2020.103416.

[5] Song K,Yan F, Ding T, Gao L and Lu S B. Computational Materials Science 174.(2020):. doi:10.1016/j.commatsci.2019.109472.

[6] Ma J, Cheng J C P , Xu Z R, Chen K Y, Lin C Q and Jiang F F. Journal of Cleaner Production 274:prepublish(2020):. doi:10.1016/j.jclepro.2020.122835.

[7] Scaringi G, Fan X M, Xu Q, Liu C, Ouyang C J, Domènech G, Yang F and Dai L X. Landslides 15.7(2018):. doi:10.1007/s10346-018-0953-9.

[8] Liu Y, Zhang D, Wang G Y, Liu C and Zhang Y. Journal of Zhejiang University-SCIENCE A 20.10(2019):. doi:10.1631/jzus.A1900292.