Deep Neural Networks for Prediction of Undrained Shear Strength of Clays

Vahid Asghari[1], Andy Y.F. Leung[2], and Mark S.C. Hsu[3]

[1]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong.
E-mail: vahid.asghari@connect.polyu.hk
[2]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong.
E-mail: andy.yf.leung@polyu.edu.hk
[3]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong.
E-mail: mark.hsu@polyu.edu.hk

Abstract: This study focuses on application of novel machine learning approaches such as deep neural networks in geotechnics, to improve the understanding on the relationships between various soil parameters and index properties. A general approach is developed using deep neural networks, to predict the undrained shear strength (s_u) of clays, based on soil data from many sites in Sweden and Finland, available in the TC304 database. The liquid limit, plastic limit, water content, vertical effective stress and preconsolidation stress are used as regressors in this model for prediction of s_u. The model has been constructed, numerous times, on a random portion of the dataset as the training set and validated on the remaining part of the data as the test set. The results of this deep neural networks model, and comparisons with the performance of other traditional methods are presented. It is shown that deep neural networks outperform these traditional approaches and can be a useful tool for future research and application in geotechnics. Although this study has been conducted on a limited data set, with sufficient amount of data, a new model can potentially be built based on the proposed approach, for prediction of soil properties in different parts of the world with reasonable reliability.

Keywords: Undrained shear strength; clays; deep neural networks; prediction; machine learning.

1 Introduction

Various sources of uncertainty are involved in geotechnical engineering, including natural variability of the geomaterials, measurement error and model uncertainty, etc. These propagate to the uncertainty associated with determination of design parameters and ultimately the system performance, often making it difficult to base the judgment purely on deterministic estimates of physical models. Various empirical, semi-empirical or probabilistic approaches have been proposed to tackle these issues, while machine learning techniques have developed rapidly in the past decades and may provide new perspective in this regard. Earlier applications of machine learning in geotechnics include those for site investigation (Zhou and Wu 1994; Juang et al. 2001), modelling of soil stress-strain response (Ellis et al. 1995; Zhu et al. 1998; Feng et al. 2004), foundation analyses (Goh 1996; Abu-Kiefa 1998; Shahin et al. 2002; Pal and Deswal 2008), and tunnelling problems (Shi et al. 1998; Shi 2000; Mahdevari et al. 2014), etc. Over the years, the development of new methods and accumulation of geotechnical data have accelerated, and it is therefore timely to consider the application of advanced machine learning techniques in geotechnics.

Deep neural networks (DNN) are a member of Deep Learning family which enables the formulation of strong regressor or classifier models. These models are being used daily in image processing, voice recognition, and similar technologies. DNN consists of many hidden layers between the input variables and the output variables (Park et al. 2009). Their applications, especially in computer vision, consist of different types of layers such as dropout and/or convolutional layers. Essentially, the main differences between traditional neural networks and DNN are the number and type of hidden layers. However, the training and structuring of DNNs are complex and there is still no one-size-fits-all approach that is applicable to all conditions. In this regard, this paper presents an approach to facilitate the DNN training and assessment of the created model. A DNN model has been formulated for clays tested in Sweden and Finland, as an example to show the potential of DNNs being used in prediction and obtaining correlations in geotechnical engineering.

The main purpose of this study is to present a new approach to predict the undrained shear strengths (s_u) of soils, based on data of tests in Finland and Sweden. Results of this study show the merits of machine learning algorithms. The results also indicate that deep neural networks outperform other traditional models. In the subsequent sections, the methodology for creating an authentic and valid deep neural network will be discussed. The results and discussion about the performance of the machine learning algorithm are then presented, together with its comparison with traditional methods. Recommendations on future research directions are also discussed.

Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)
Editors: Jianye Ching, Dian-Qing Li and Jie Zhang

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0_MS2-4-cd

2 Methodology

The datasets investigated in this study are referred to as F-CLAY/7/216 and S-CLAY/7/168 (D'Ignazio et al. 2016), both accessible from the TC304 database. The first clay dataset has been compiled from field vane tests conducted in 24 different test sites in Finland. Each one of the 216 samples in this dataset includes information on undrained shear strength, vertical effective stress (σ'_v), preconsolidation stress (σ'_p), liquid limit (LL) and plastic limit (PL). The second clay dataset contains 168 samples of field vane tests conducted in 12 sites in Sweden and 7 sites in Norway, and was part of the CLAY/10/7490 database (Ching and Phoon 2014a).

The two datasets of Scandinavian soils are combined and studied together herein using the DNN. The creation and design of prediction or classification models using DNN are usually time-consuming. This is mainly because of the associated computational demands, and the lack of a clear way to structure these models. A machine learning model is valid when it can predict with acceptable error both on the training set and the test set. It should also be able to reconstruct and not be limited to a specific proportion of the data. In addition, since there are many possible models which could be created for DNN, comparison between different models is necessary. Taking these aspects into consideration, five main steps are recommended herein to pave the way toward creating an authentic and valid model. The aforementioned steps are as follows: 1) data preprocessing, 2) training curve analysis, 3) structuring the neural network model, 4) assessment and solution of overfitting problem, and 5) analysis of the created model. The results and interpretation for each of these steps will be briefly discussed in the following sections.

2.1 Data preprocessing

Optimization of a cost function is the most important step in training a neural network. The speed and required time for convergence of the cost function optimization depend on the adopted optimization algorithm and the features of the data. Datasets with different scales, distributions, and dimensions would significantly affect the optimization time. They can also affect the effectiveness of the optimizer, occasionally hindering the algorithm from reaching the optimum point. In addition, existence of outliers would affect the training if this aspect is not given proper consideration. To solve this problem, standardization or normalization of the data are important. By equalizing the range and distribution of input variables, these can help the optimizer to converge to the optimum point more efficiently, and they also mitigate the presence of outliers. Standardization converts the mean and standard deviation of the data to zero and one, respectively:

$$f_s(v) = \frac{v - \mu_i}{\sigma_i} \tag{1}$$

where f_s is the standardizer function, v is a value from a series of input variable i into the model, μ_i is the mean of the input variable i, and σ_i is the standard deviation of the input variable i. Equation (2) shows the normalization function f_n, which sets the maximum and minimum of a series of data to 1 and 0, respectively.

$$f_n(v) = \frac{v - \min(i)}{\max(i) - \min(i)} \tag{2}$$

where max and min are the maximum and minimum of the corresponding input variable series. Normalization would lessen the information given by the data series in case of presence of outliers or deviants from the mean of the data series. If the data is uniformly distributed, normalization could be other choice of feature scaling (Ng 2016).The dataset used in this study consist of two types of parameters:

1. LL, PL, and w (moisture content) which are dimensionless and presented in percentages.
2. σ'_p and σ'_v which are presented in kPa.

As an example, Figure 1 shows the liquid limit and vertical effective stress of samples in the dataset. As shown in the figures, the data points are not distributed uniformly. Therefore, normalization would not be a proper choice for preprocessing of this input variable. Standardization has been used for all of the input variables in this study. Since the purpose of this study is to predict the undrained shear strength as the output variable, it has not been transformed and linear activation function has been used for the output layer.

As there are numerous options for structure of the DNN, it would be more practical to divide the dataset into three subsets: 1) training set which will be used for training the model; 2) cross-validation set which will be used to find the best DNN structure and to compare the models with each other; and 3) test set which will be used for reporting the performance of the model. The test set and cross-validation set will not be included in the training. In this study, the data has been divided by 60%, 20% and 20% for the aforementioned subsets respectively. However, with larger datasets, these proportions for dividing the dataset can be adjusted. It is noteworthy that removal of outliers from the dataset will improve the performance of the model; however, to show the effectiveness of the DNN, the model has been created on the entire dataset in this study.

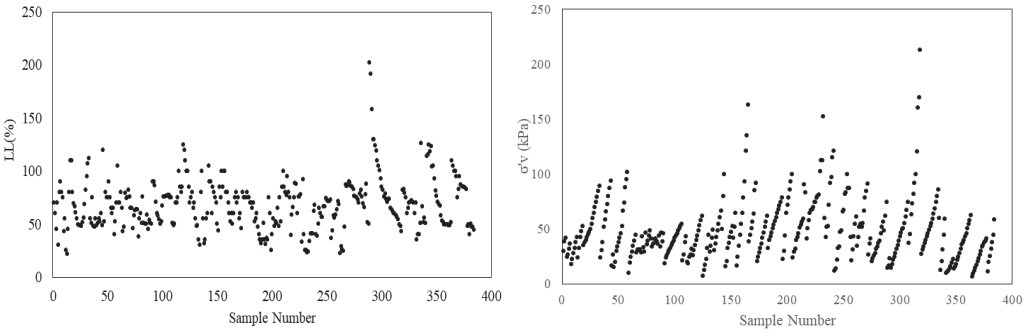

Figure 1. (left) Liquid limit of samples; (right) vertical effective stress of samples.

2.2 Training curve analysis

Training curve analysis provides valuable information about the data and saves time and efforts for the analysts. By interpreting the training curve, the analysts may decide whether to abandon the project, gather more data, or make the structure of the model more complex. In this process, the training error and cross-validation error will be evaluated and reported within j steps. At each step, a portion that equals to i/j of the training set is chosen for training. Figure 2 depicts the results of the training curve in 10 steps.

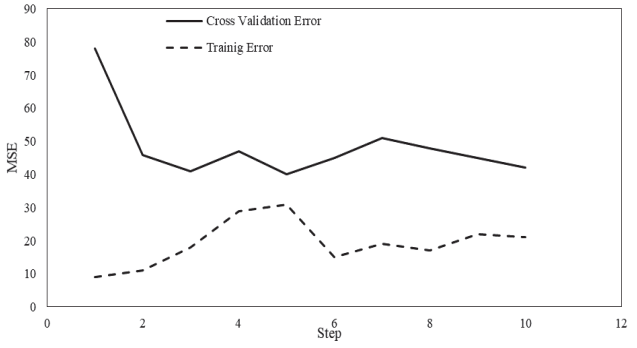

Figure 2. Training curve results.

As a starting point, two hidden layers, each containing 10 nodes, have been used for training curve analysis. The structure of the DNN is not of high importance in this step, while interpretation of the figure provides the following observations:

1. Although the cross-validation error and training error lines are jagged, they are about to converge and more samples would lead to a model with less error.
2. There is a separation between the cross-validation error line and the training error line, which indicates the problem of overfitting. This issue needs to be resolved.
3. The simplest prediction model is the average of the output variable. The mean of squared error of this model (i.e. the variance) is the maximum error for any other prediction model. As shown in Figure 2, the mean of squared error (MSE) of the predictions is around 30, which is significantly less than the variance of the output variable. This means that with a proper neural network structure, a relatively reliable model can be created.

2.3 Structuring the neural network

DNN structure has some characteristics that significantly affect the model performance. Activation functions, the number and type of hidden layers and nodes, optimization algorithm, and loss function are the main parts of the structure of a DNN. Ng (2016) provided guidelines and tips for choosing activation functions. The *Tanh* function has been used as the activation function for the input layer to map the nonlinearity on the input and output of the model. The *linear* activation function has been chosen for the output layer as this model is a regression model. The *Relu* activation function, also, was chosen as the activation function for the hidden layers. Although other nonlinear activation function may be used for the hidden layers, *relu* has been shown to enhance efficiency of the

optimization (Ng 2016). The Adam optimizer is one of the efficient algorithms that has been used for training in this study. The MSE of the predictions has been monitored and will be minimized in training. There is no certain way to find the optimal number of hidden layers and neurons. After a series of trial-and-error attempts with numerous models, the model with three layers of 15, 10, and 5 nodes was proposed, respectively. With the presented information regarding the structure of the DNN, it is possible to replicate the results of this study. However, since the training set, cross-validation set, and test set have been chosen randomly, the final results may be slightly different after each training, but this should not significantly alter the assessment and validity of the formulated DNN models.

2.4 Assessment and solution of overfitting problem

Overfitting is a common issue in any type of statistical or machine learning problem. This refers to an acceptable performance (small error) of the fitted model in training set, but weak performance (large error) on the test set. The previously created model (Figure 2) for prediction of undrained shear strength suffers from overfitting problem, as the MSE of the training set and cross-validation set differ significantly. Therefore, proper actions should be taken to ameliorate this issue.

Dropout layers and regularization are two main ways of solving the overfitting problem in DNN. Dropout layers activate a node with a determined probability, which was proposed to be set to 0.5 (Srivastava et al. 2014). It is mainly used in machine vision problems with large number of inputs. Regularization deviates the hidden coefficients, in the hidden layers and nodes, in such a manner that the cross-validation error and training error would be equalized.

Figure 3 depicts the regularization parameter analysis. Starting from a value of 10^{-7}, the regularization parameter is increased in each step by 3 times compared to the previous value. The step corresponding to the minimum cross-validation error indicates the optimal choice for the regularization parameter. Therefore, 2.0×10^{-4} has been chosen as the regularization parameter. Accordingly, the MSE of the three subsets (training set, cross validation set, and test set) are found to be 19.8, 24.1 and 21.1, respectively. Since the MSE in these sets are not significantly different, the model is considered to be valid and acceptable.

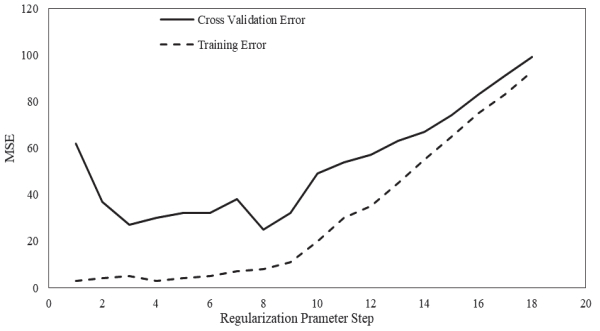

Figure 3. Regularization parameter analysis.

2.5 Model performance

Figure 4 shows the predicted values of s_u versus actual values. This figure shows that the performance of the created model using DNN is also reasonable. The R^2 is found to be 0.86, with MSE of 28.5 and mean of absolute errors (MAE) of 4.0.

Although DNN models are often considered as 'black-box' algorithms, researchers (Ribeiro et al. 2016; Lundberg and Lee 2017) have proposed various indicators for assessments of interpretability, such as the 'SHAP' (SHapley Additive exPlanations) values, which represent the features' importance for deep models using game theory. SHAP represents the importance of each feature in every prediction considering the 'contributors' are the features and the game is predicting the output variable (Lundberg and Lee 2017).

Figure 5 shows the average of relative importance of features in all predictions. In the current model, σ'_p and σ'_v are the most important features, while the importance of other features is not significant in comparison to σ'_p.

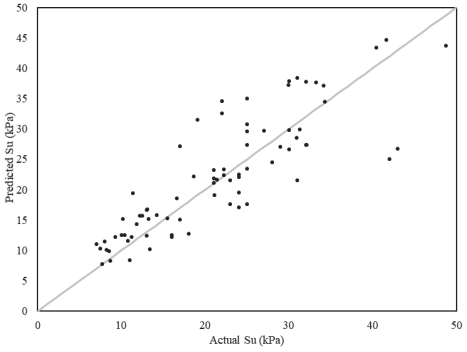

Figure 4. Predicted versus actual undrained shear strengths.

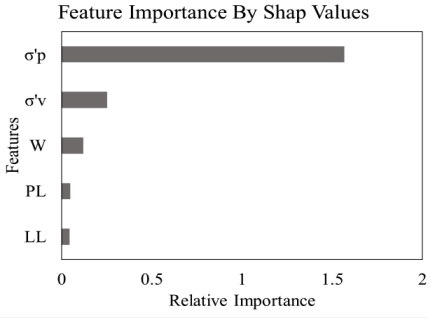

Figure 5. Feature importance of designed DNN model.

3 Results and Discussion

Correlation relationships are commonly used in the geotechnical community for estimation of soil properties based on other measured parameters and indices. For example, the datasets of Finnish and Swedish clays investigated in this study were also analyzed by D'Ignazio et al. (2016), and Ching and Phoon (2014a; 2014b), who proposed a number of relationships to correlate the s_u with other field and laboratory tests of clays. To compare the performance of the DNN proposed in this study, the dataset is also analyzed using other traditional techniques, including linear regression and artificial neural networks (ANN). Table compares the performance of the different techniques. Based on these results, DNN outperform the other two methods. In fact, it also outperforms the models presented in D'Ignazio et al. (2016).

Table 1. Comparison of performance of the proposed and previous models.

Model	MSE	MAE	MAPE	R^2	Comments
$\ln(\frac{S_u}{\sigma_v'}) = -0.87 + 0.24\text{LI}$	322.35	12.48	74.91%	0.45	Ching and Phoon (2014b)
$\ln(\frac{S_u}{\sigma_v'}) = -1.47 + \ln(\text{OCR})$	58.38	5.28	27.63%	0.75	Ching and Phoon (2014b)
$\frac{S_u}{\sigma_v'} = 0.23\ \text{OCR}^{0.8}$	57.93	5.38	27.77%	0.73	Ching and Phoon (2014a)
Linear Regression	41.72	4.71	31.36%	0.73	
ANN	35.67	4.55	31.70%	0.78	
DNN model	28.5	4.01	23.40%	0.86	The proposed model

As a member of machine learning algorithms, DNN are one of the most powerful tools for predictions and classification problems. However, they have not been given sufficient attention in geotechnical engineering. Not only are the results of DNN reproducible, the approach can also be applied to other areas of geotechnics, as a new way to obtain relationships between various index properties and geotechnical parameters, and to potentially

reveal hidden mechanisms that cause discrepancies between predicted and actual field response in soil-structure interaction problems. Previous works on geotechnical correlations were often based on fundamental soil mechanics concepts, and supplemented by empirical evidence through regression. On the contrary, the current DNN model (and many other machine learning approaches) is solely based on statistical or 'numerical' training of the existing dataset, without explicitly incorporating the mechanical aspects of the problem. While the results in Table 1 show that this model is not inferior to previous methods, further developments of DNN, with its applications to larger worldwide datasets, will hopefully lead to better insights into the physical phenomena of geotechnical correlations, besides improved prediction capabilities for the models.

Only three layers are included in the DNN for this study, as the number of data points is only about 400. Nonetheless, the proposed framework, with the five main steps for preprocessing, model structuring and assessments, can form a solid basis to analyze larger datasets, and to train deeper networks with convolutional layers and dropout layers to further enhance the performance.

Acknowledgments

The authors would like to thank the members of the TC304 Committee on Engineering Practice of Risk Assessment & Management of the International Society of Soil Mechanics and Geotechnical Engineering for developing the database 304dB used in this study and making it available for scientific inquiry. We also wish to thank M. D'Ignazio, T.T. Länsivaara, J. Ching and K.K. Phoon for contributing this database to the TC304 compendium of databases. The study is partially supported by the Research Grants Council of the Hong Kong Special Administrative Region (Project No.15217517).

References

Abu-Kiefa, M.A. (1998). General regression neural networks for driven piles in cohesionless soils. *Journal of Geotechnical and Geoenvironmental Engineering*, 124(12), 1177-1185.

Ching, J. and Phoon, K.K. (2014a). Transformations and correlations among some clay parameters — the global database. *Canadian Geotechnical Journal*, 51(6), 663–685.

Ching, J. and Phoon, K.K. (2014b). Correlations among some clay parameters — the multivariate distribution. *Canadian Geotechnical Journal*, 51(6), 686–704.

D'Ignazio, M., Phoon, K.K., Tan, S.A., and Länsivaara, T.T. (2016). Correlations for undrained shear strength of Finnish soft clays. *Canadian Geotechnical Journal*, 53(10), 1628–1645.

Ellis, G.W., Yao, C., Zhao, R., and Penumadu, D. (1995). Stress-strain modelling of sands using artificial neural networks. *Journal of Geotechnical Engineering*, 121(5), 429-435.

Feng, X.T., Zhao, H., and Li, S. (2004). A new displacement back analysis to identify mechanical geo-material parameters based on hybrid intelligent methodology. *International Journal for Numerical and Analytical Methods in Geomechanics*, 28(11), 1141-1165.

Goh, A.T.C. (1996). Pile driving records reanalyzed using neural networks. *Journal of Geotechnical Engineering*, 122(6), 492-495.

Juang, C.H., Jiang, T., and Christopher, R.A. (2001). Three-dimensional site characterization: neural network approach. *Géotechnique*, 51(9), 799–809.

Lundberg, S.M. and Lee S.I. (2017). A unified approach to interpreting model predictions. *Proc. 31st Conference on Neural Information Processing Systems*.

Mahdevari, S., Shahriar, K., Yagiz, S., and Shirazi, M.A. (2014). A support vector regression model for predicting tunnel boring machine penetration rates. *International Journal of Rock Mechanics and Mining Sciences*, 72, 214-229.

Ng, A. (2016). *Machine Learning Yearning*.

Pal, M. and Deswal, S. (2008). Modeling pile capacity using support vector machines and generalized regression neural network. *Journal of Geotechnical and Geoenvironmental Engineering*, 134(7), 1021-1024.

Park, C., Lee, S.B., and An, K.H. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 20(2), 1-127.

Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. Available at: http://arxiv.org/abs/1602.04938.

Shahin, M.A., Maier, H.R., and Jaksa, M.B. (2002). Predicting settlements of shallow foundations using neural networks. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(9), 785–793.

Shi, J.J. (2000). Reducing prediction error by transforming input data for neural networks. *Journal of Computing in Civil Engineering*, 14(2), 109-116.

Shi, J.J., Ortigao, J.A.R., and Bai, J. (1998). Modular neural networks for predicting settlement during tunneling. *Journal of Geotechnical and Geoenvironmental Engineering*, 124(5), 389-395.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 520–525.

Zhou, Y. and Wu, X. (1994). Use of neural networks in the analysis and interpretation of site investigation data. *Computer and Geotechnics*, 16, 105-122.

Zhu, J. H., Zaman, M.M., and Anderson, S.A. (1998). Modeling of soil behavior with a recurrent neural network. *Canadian Geotechnical Journal*, 35(5), 858-872.