Catena 166 (2018) 181–191

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena

Prediction of shear strength of soft soil using machine learning methods

Binh Thai Pham[a], Le Hoang Son[b], Tuan-Anh Hoang[b], Duc-Manh Nguyen[c], Dieu Tien Bui[d,e,*]

[a] Geotechnical Engineering and Artificial Intelligence research group (GEOAI), University of Transport Technology, Ha Noi, Viet Nam
[b] VNU University of Science, Vietnam National University, Viet Nam
[c] Department of Geotechnical Engineering, University of Transport and Communication, Ha Noi, Viet Nam
[d] Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[e] Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

ARTICLE INFO

Keywords:
Artificial neural networks
Machine learning
Particle swarm optimization
Strength of soft soils

ABSTRACT

Shear strength of the soil is an important engineering parameter used in the design and audit of geo-technical structures. In this research, we aim to investigate and compare the performance of four machine learning methods, Particle Swarm Optimization - Adaptive Network based Fuzzy Inference System (PANFIS), Genetic Algorithm - Adaptive Network based Fuzzy Inference System (GANFIS), Support Vector Regression (SVR), and Artificial Neural Networks (ANN), for predicting the strength of soft soils. For this purpose, case studies of 188 plastic clay soil samples collected from two major projects, Nhat Tan and Cua Dai bridges in Viet Nam have been used for generating training and testing datasets for constructing and validating the models. Validation and comparison of the models have been carried out using RMSE, and R. The results show that the PANFIS has the highest prediction capability (RMSE = 0.038 and R = 0.601), followed by the GANFIS (RMSE = 0.04 and R = 0.569), SVR (RMSE = 0.044 and R = 0.549), and ANN (RMSE = 0.059 and R = 0.49). It can be concluded that out of four models the PANFIS indicates as a promising technique for prediction of the strength of soft soils.

1. Introduction

In geotechnical engineering, the shear strength of the soil is an important engineering parameter which is certainly used in the design and audit of many geo-environmental and geo-technical structures i.e. road foundations and pavements, earth dams, and retaining walls (Vanapalli and Fredlund, 2000). It is determined by two important parameters to determine the shear strength, internal friction angle and unit cohesion (Das and Sobhan, 2013), and affected by several factors namely plastic index (*PI*), liquid limit (*LL*), moisture content (W), clay content (CC), etc. (Das and Sobhan, 2013; Kaya, 2009). It increases together with the approximate volume of grouted zone for treated samples soil with cement grout in the study about effects of the permeation cement grout with fly ash on the sandy soil skeleton (Ali and Yousuf, 2016; Vanapalli and Fredlund, 2000).

Many studies have been carried out for the prediction of the shear strength of soft soils. Motaghedi and Eslami (2014) proposed an analytical approach for C, ϕϕ prediction using all quantities, q_c, u_2, and f_s considering bearing capacity mechanism of failure at cone tip and direct shear failure along the penetrometer sleeve. McGann et al. (2015) used a multiple linear regression to develop a Christchurch-specific empirical correlation for predicting soil shear wave velocities (Vs) from cone penetration test (CPT) data. Azari et al. (2014) studied the effects of shear strength variation in the disturbed zone on the time-dependent behavior of soft soil deposits improved with vertical drains and preloading. Griffiths et al. (2016) used equivalent linear and nonlinear 1D site response analyses for the well-known Treasure Island site to demonstrate challenges associated with accurately modeling large shear strains, and subsequent surface response, at soft soil sites. Oliveira et al. (2017) investigated constitutive models to simulate the creep behavior of a soft soil in its natural state or chemically stabilized state. It has been inferred from those studies that a well-established mathematical model should be constructed in order to achieve high accuracy of prediction.

In recent decades, **machine learning or artificial intelligent methods** have been applied widely for generating such the prediction models of material properties (Shahin et al., 2009; Pham et al., 2017; Pourghasemi and Rahmati, 2018; Shirzadi et al., 2017). Samui (2008) applied Support Vector Regression (SVR) for predicting the friction capacity of driven piles in clay soils. Behavior prediction of shallow foundations was also carried out using the Artificial Neural Network (ANN) in several studies including bearing capacity (Kuo et al., 2009; Padmini et al., 2008). Chou et al. (2016) used data mining including linear regression, classification and regression tree (CART) analysis, a

* Corresponding author at: Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Viet Nam.
E-mail address: buitiendieu@tdt.edu.vn (D. Tien Bui).

https://doi.org/10.1016/j.catena.2018.04.004
Received 2 February 2018; Received in revised form 9 March 2018; Accepted 3 April 2018

generalized linear (GENLIN) model, chi-squared automatic interaction detection (CHAID), ANN, and SVR to identify factors influencing shear strength and to predict the peak friction angle of FRS. In prediction of shear strength of soil, there are several studies. Das et al. (2011) studied the potential of the SVM and ANN for prediction of the residual strength of soil. Kanungo et al. (2014) compared the ANN and CART techniques for predicting the shear strength parameters. Kiran et al. (2016) applied Probabilistic Neural Network (PNN) to predict the shear strength parameters of soil, viz., cohesion "c" and internal friction angle "φ" from water content (w), Plasticity Index (PI), Dry Density (DD), Gravel % (GP), Sand % (SP), Silt % (STP), and Clay % (CP) of soil. Prediction of residual strength of clay based on a new prediction model namely functional network (FN) has been investigated in Khan et al. (2016). In general, the common conclusion from the aforementioned works is that machine learning methods are efficient for prediction of shear strength of soft soils (Moavenian et al., 2016).

The recent development of machine learning and optimization have resulted in some new promising soft computing methods i.e. Particle Swarm Optimization - Adaptive Network based Fuzzy Inference System (PANFIS), Genetic Algorithm - Adaptive Network based Fuzzy Inference System (GANFIS). PANFIS and GANFIS are state-of-the art methods that were formed by integrating meta-heuristic optimization algorithms and neural fuzzy models. They have proven as the powerful tools in predicting various environmental problems such as flood (Bui et al., 2016a), forest fire (Bui et al., 2017), displacement of hydropower dam (Bui et al., 2016b), and landslide (Chen et al., 2017). On the other hand, SVR and ANN are popular and efficient methods used in the shear strength modeling. However, investigation and comparison of these methods with popular machine learning methods i.e. Support Vector Regression (SVR) and Artificial Neural Networks (ANN) for the prediction of the shear strength of soft soils have not been carried out.

In this study, we expand the body of knowledge thought investigating and comparing the prediction performance of PANFIS, GANFIS, SVR, and ANN for the prediction of shear strength of soft soil. The comparison of such the machine learning methods is significant for determination of an effective prediction model that can be used in practical scenarios of shear strength of soft soils.

The rest of the paper is organized as follows. Section 2 presents the study sites and dataset description. Section 3 gives the background of the models including PANFIS, GANFIS, SVR, and ANN. Sections 4 and 5 demonstrate the results and discussion. Lastly, Section 5 draws conclusions and suggests further studies. It is noted that MatlabR2014b and Weka 3.8.1 were used for dataset generation and modeling.

2. Study site and data

2.1. Description of the study site

In this research, plastic clay soil samples from two bridge construction projects, the Nhat Tan Bridge (Ha Noi City) and the Cua Dai Bridge (Quang Nam City) in Vietnam were used as a case study. The Nhat Tan Bridge is located in about Latitude 20°50′30″N and Longitude 106°41′37″E, whereas the Cua Dai Bridge is located in Latitude 15°53′25″E and Longitude 108°20′42″E (Red points on the map in Fig. 1). The main beam system of the Nhat Tan Bridge was designed and constructed using cable-stayed structure with five diamond towers and six spans. The whole length of the Cua Dai Bridge is 18.3 km, and the bridge part on the river is 1.481 km.

2.2. Data

A total of 188 samples from the two bridge projects were collected and used for generating the datasets for modeling. In this prediction problem, the shear strength is the output variable whereas the input variables are moisture content, clay content, liquid limit, plastic limit, plastic index, and consistency index.

2.2.1. Shear strength

"Shear strength (τ) of a soil mass is the internal resistance per unit area that the soil mass can offer to resist failure and sliding along any plane inside it" (Das and Sobhan, 2013). It is an important factor in analyzing the soil stability problems including slope stability, lateral pressure on earth-retaining structures, and bearing capacity. The failure of a soil mass is not due to either shear stress or maximum normal alone and because of a critical combination of shearing stress and normal stress (Das and Sobhan, 2013). Therefore, the functional relationship between shear stress and normal stress on a failure plane of a soil mass can be presented as follows:

$$\tau = f(\sigma) = \sigma \tan\varphi + c, \tag{1}$$

where σ (kg/cm^3) is the normal stress on the failure plane, φ is the angle of internal friction, and c (kg/cm^3) is the cohesion (Das and Sobhan, 2013).

In the laboratory, the parameters of shear strength (c, φ) can be determined using different experiments namely direct shear test, triaxial test, and torsional ring shear test (Das and Sobhan, 2013; Whitlow, 1990). In general, the determination of these parameters for calculating the shear strength of a soil mass is relatively complicated and costly. In this study, suppose $\sigma = 1kg/cm^2$, the shear strength was calculated using the parameters (c, φ) determined by direct shear test from 188 plastic clayed soil samples as follows:

$$\tau = \tan\varphi + c,\ \sigma = 1kg/cm^2. \tag{2}$$

Data of shear strength of 188 plastic clayed soil samples is shown in Fig. 2. It shows that τ values differ from 0.104 to 0.301 (kg/cm^3), the mean value is 0.197 (kg/cm^3), and the standard deviation value is 0.047 (kg/cm^3).

2.2.2. Moisture content

"Moisture content (ω) is also referred to as water content and is defined as the ratio of the weight of water to the weight of solids in a given volume of soil" (Das and Sobhan, 2013; Whitlow, 1990). Moisture content affects the shear strength of soil as it reduces the cohesive forces between soil solids, and even causes the saturation of soils. As the moisture content increases the shear strength of soils reduces (Sharma and Bora, 2003). Thus, moisture content was taken into account as an affecting factor for predicting of the shear strength of soils in this study. Moisture content is determined in laboratory using an oven drying method or field test using alcohol burning method.

Moisture content can be calculated using following equation (Das and Sobhan, 2013; Whitlow, 1990):

$$\omega(\%) = \frac{W_\omega}{W_s} \times 100 = \frac{m_\omega g}{m_s g} \times 100, \tag{3}$$

where W_ω is the weight of water of soil sample, W_s is the weight of solids of soil sample, m_ω is the mass of water of soil sample, m_s is the mass of the solids of soil sample, and g is the gravity acceleration ($g = 9.81\ m/s^2$). In this study, moisture content test was carried out in the laboratory, and the moisture content values of 188 samples are shown in Fig. 3a. It shows that the moisture content values vary from 24.19 to 141.83 (%), the mean value is 56.1 (%), and the standard deviation value is 19.1 (%).

2.2.3. Clay content

Clays are classified as the soil solids smaller than 0.002 mm in size. In several cases, the soil solids between 0.002 and 0.005 mm in size are also considered as clays (Das and Sobhan, 2013). Clay content (μ) was considered as an affecting factor to the shear strength of soils as it develops the plasticity of soils, and as the clay content increases the shear strength of soils reduces when soils are mixed with a limited amount of water. Clay content can be determined in the laboratory using grain size distribution analyzing test through following equation (Das and Sobhan, 2013):