Multivariate Adaptive Regression Spline and Least Square Support Vector Machine for Prediction of Undrained Shear Strength of Clay

Pijush Samui, Vellore Institute of Technology University, India
Pradeep Kurup, University of Massachusetts Lowell, USA

ABSTRACT

This study adopts Multivariate Adaptive Regression Spline (MARS) and Least Square Support Vector Machine (LSSVM) for prediction of undrained shear strength ($s_u$) of clay based on Cone Penetration Test (CPT) data. Corrected cone resistance ($q_c$), vertical total stress ($\sigma_v$), hydrostatic pore pressure ($u_h$), pore water pressure at the cone tip ($u_1$), and pore water pressure just above the cone base ($u_2$) are used as input parameters for building the MARS and LSSVM models. The developed MARS and LSSVM models give simple equations for prediction of $s_u$. A comparative study between MARS and LSSVM is presented. The results confirm that the developed MARS and LSSVM models are robust for prediction of $s_u$.

Keywords: Clay, Cone Penetration Test (CPT), Least Square Support Vector Machine (LSSVM), Multivariate Adaptive Regression Spline, Undrained Shear Strength

INTRODUCTION

The determination of undrained shear strength ($s_u$) of clay is an important task for the analysis and design of geotechnical systems. It has been used to determine the effect of an earthquake on soil, stability of clay slope, etc. Geotechnical engineers routinely use empirical correlations to estimate $s_u$ from Cone Penetration Test (CPT) data (Lunne et al., 1997). The following formula may be used to determine $s_u$ of clay (Lunne et al., 1997).

$$q_c = N_k S_u + \sigma_v$$  \hspace{1cm} (1)

where $q_c$ is the cone tip resistance, $N_k$ is the empirical cone factor, and $\sigma_v$ is the total vertical stress at the depth of penetration. However, there is no single value of $N_k$ for all type of clays, penetrometers and test conditions (Amar et al.,...
There are numerous empirical correlations between $q_c$ and $s_u$ reported in literature (Lunne et al., 1976; Koutsoftas & Fischer, 1976; Stark & Delashaw, 1990).

This article adopts Multivariate Adaptive Regression Spline (MARS) and Least Square Support Vector Machine (LSSVM) for developing correlations between $s_u$ and CPT test data. MARS is a flexible, more accurate, and faster simulation method (Friedman, 1991; Salford Systems, 2001). It has been successfully used to solve different problems (MacLean & Mix, 1991; Veaux et al., 1993; Ekman & Kubin, 1999; Prasad & Iverson, 2000; Jin et al., 2000; Ko et al., 2004; Sharda et al., 2008; Okine et al., 2003, 2009). LSSVM is proposed by Suykens and Vandewalle (1999). In LSSVM, the training is expressed in terms of solving a set of linear equations in the dual space instead of quadratic programming. Researchers have successfully used to solve different problems in engineering (Pahasa & Ngamroo, 2011; Deng & Yeh, 2010; Huang et al., 2009; Bin et al., 2008).

This article uses the database obtained from Sandven (1990) and Chen (1994) that included data from clay sites located in several countries around the world. The dataset contains information about the corrected cone resistance ($q_t$), vertical total stress ($\sigma_v$), hydrostatic pore pressure ($u_0$), pore pressure at the cone tip ($u_1$), pore pressure just above the cone base ($u_2$) and $s_u$. The reference values of $s_u$ were estimated from isotropically and anisotropically consolidated undrained triaxial compression (CIUC and CAUC) tests and in situ field vane shear tests.

A comparative study has been done between the developed LSSVM and MARS. The organization of the work is as follows. The details of MARS are proposed in the upcoming section. Then the LSSVM technique is described. The results from the developed MARS and LSSVM are presented afterwards. The conclusions are given in the last section.

### DETAILS OF MARS

MARS is a flexible modeling method for high-dimensional data (Friedman, 1991). A brief overview of the MARS model will be given here. MARS uses the following equation for prediction of output $y$.

$$y = \alpha_0 + \sum_{m=1}^{M} \alpha_m B_m(x)$$  \hspace{1cm} (1)

Where $B_m(x)$ is basis function, $\alpha_m$ is the coefficient of $B_m(x)$, $x$ is the input variables, $M$ is the number of basis functions and $\alpha_0$ is constant. In this study, the input parameters are $q_t$, $\sigma_v$, $u_0$, $u_1$, $u_2$. The output of the MARS model is $s_u$. So, $x = [q_t, \sigma_v, u_0, u_1, u_2]$ and $y = [s_u]$.

MARS adopts the following truncated spline function as basis function (Sekulic & Kowalski, 1992).

$$b^-(x-t) = \begin{cases} (t-x)_+^q & \text{if } x < t \\ 0 & \text{otherwise} \end{cases}$$ \hspace{1cm} (2)

$$b^+(x-t) = \begin{cases} (x-t)_+^q & \text{if } x > t \\ 0 & \text{otherwise} \end{cases}$$ \hspace{1cm} (3)

Where $b^-(x-t)$ and $b^+(x-t)$ are spline function, $t$ is knot location, and $q$ is the power.

MARS uses the two steps to construct the final model. In the first step, basis functions are introduced to define equation 1. Overfitting can occur due to many basis functions. This step is called forward stepwise procedure. In backward stepwise procedure, redundant basis are removed based on generalized cross-validation (GCV) (Sekulic & Kowalski, 1992). The GCV criterion is defined as follows:
Modeling and Simulation of a Stand-Alone Hydrogen Photovoltaic Fuel Cell Hybrid System

Discrete Artificial Bee Colony Optimization Algorithm for Financial Classification Problems
[www.igi-global.com/article/discrete-artificial-bee-colony-optimization/52789?camid=4v1a](www.igi-global.com/article/discrete-artificial-bee-colony-optimization/52789?camid=4v1a)