Int. J. Advance. Soft Comput. Appl., Vol. 3, No. 1, March 2011
ISSN 2074-8523; Copyright © ICSRS Publication, 2011
www.i-csrs.org

Determination of Compression Index For Marine Clay: A Least Square Support Vector Machine Approach

Pijush Samui, Sarat Das, Dookie Kim and Gil Lim Yoon

Associate Professor, Centre for Disaster Mitigation and Management, VIT Univeristy, Vellore-632014, India.
email: pijush.phd@gmail.com,
Associate Professor, Department of Civil Engineering, National Institute of Technology Rourkela, India- 769 008.
email:saratdas@rediffmail.com
Associate Professor, Department of Civil Engineering, Kunsan National University, Kunsan, Jeonbuk, South Korea.
email:kim2kie@chol.com
Senior researcher, Coastal harbor engineering Lab,Korea Ocean Research & Development Institute (KORDI),Ansan 425-744 Korea..
email: glyoon@kordi.re.kr

Abstract

This article employs Least Square Support Vector Machine (LSSVM) for determination of Compression Index (C_c) of marine clay in east coast of Korea. This study uses LSSVM as a regression tool. In LSSVM, the regression equation is obtained as the solution to a linear system instead of a quadratic programming (QP) problem. The input parameters of LSSVM are natural water content (ω_n), liquid limit (ω_l), initial void ratio (e_0), and plasticity index (PI). Equations have been also developed for the prediction of C_c of marine clay. The comparison between the developed LSSVM and the regression models shows that the developed LSSVM models perform better than the regression models. This article shows that the developed LSSVM can be used to predict C_c of marine clay in east coast of Korea.

Keywords: *Marine Clay; Compression Index; Least Square Support Vector Machine; Prediction.*

1 Introduction

Marine clay is highly compressible and produces high settlement. Compression index (C_c) is an important parameter to determine the settlement of marine clay. So, the determination of C_c of marine clay is an imperative task in earth science. Researchers use different correlations for determination of C_c (Lambe and Whitman, 1969; Mayne, 1980; Nakase et al., 1988; Skempton, 1944; Terzaghi and Peck, 1967; Yin, 1999). The available correlations are based on the regression analysis which uses least-square method for prediction. Least-square method is sensitive to the presence of outliers, and it performs poorly when the underlying distribution of the additive noise has a long tail (Haykin, 1999). So, the available correlations are not suitable for determination of C_c.

This study adopts Least Square Support Vector Machine (LSSVM) for determination of C_c of marine clay in east coast of Korea based on different properties {natural water content (ω_n), liquid limit (ω_1), initial void ratio (e_0), and plasticity index (PI). This study uses the database collected by Yoon et al.(2004). The dataset contains the information about ω_n, ω_1, e_0, PI and C_c of the marine clay in east coast of Korea. In LSSVM, Vapnik's ε-insensitive loss function has been replaced by a sum-squared error (SSE) cost function. LSSVM considers equality type constraints instead of inequalities as in the classic Support Vector Machine (SVM) approach (Suykens et al., 2002; Vapnik, 1995, 1998). LSSVM has been successfully applied for solving different problems in engineering (Bayler et al., 2009; Samui and Sitharam, 2008; Wang and Ye, 2004).

This study has the following aims:

- To examine the capability of LSSVM for prediction of C_c of marine clay in the east coast of Korea
- To develop equation for prediction of C_c of marine clay in the east coast of Korea
- To make a comparative study between the developed LSSVM and the regression models developed by Yoon et al.(2004)

2 Details of LSSVM

LSSVM models are an alternate formulation of SVM regression (Vapnik and Lerner, 1963) proposed by Suykens et al (2002). Consider a given training set of N data points $\{x_k, y_k\}_{k=1}^{N}$ with input data $x_k \in R^N$ and output $y_k \in r$ where R^N is

the N-dimensional vector space and r is the one-dimensional vector space. For prediction of C_c of marine clay using single marine clay parameter, $x = [\omega_n \text{ or } \omega_1 \text{ or } PI \text{ or } e_0]$ and $y = [C_c]$. For prediction of C_c of marine clay using multiple marine clay parameters, $x = [PI, \omega_1, e_0]$ and $y = [C_c]$.
In feature space LSSVM models take the form

$$y(x) = w^T \varphi(x) + b \tag{1}$$

Where the nonlinear mapping $\varphi(.)$ maps the input data into a higher dimensional feature space; $w \in R^n$; $b \in r$; w = an adjustable weight vector; b = the scalar threshold. In LSSVM for function estimation the following optimization problem is formulated:

Minimize: $\frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2$

Subject to: $y(x) = w^T \varphi(x_k) + b + e_k$, k=1,…,N. (2)

Where e_k=error variable and γ= regularization parameter. The following equation for C_c prediction has been obtained by solving the above optimization problem (Smola and Scholkopf, 1998; Vapnik, 1998).

$$C_c = y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{3}$$

Where $K(x,x_k)$ is kernel function. The radial basis function has been used as kernel function in this analysis. The radial basis function is given by

$$K(x_k, x_l) = \exp\left\{ -\frac{(x_k - x_l)(x_k - x_l)^T}{2\sigma^2} \right\} \qquad k,l=1,\ldots,N \tag{4}$$

Where σ is the width of radial basis function.

3 Details of Present Analysis

This article adopts the above LSSVM model for determination of C_c of marine clay in the east coast of Korea. To develop LSSVM, the data have been divided into the following two groups:

Training Dataset: This has been used for constructing the LSSVM. This study uses 180 data out of 257 as a training dataset

Testing Dataset: This has been used to assess the performance of the LSSVM. The remaining 77 data have been used as a testing dataset.

The data are normalized between 0 and 1. The normalization has been done by using the following equation:

$$d_{normalized} = \frac{(d - d_{\min})}{(d_{\max} - d_{\min})} \tag{5}$$

Where d=any data (input or output), d_{min}= minimum value of the entire dataset, d_{max}= maximum value of the entire dataset, and $d_{normalized}$=normalized value of the data. The design values of γ and σ has been determined by the trial and error approach during the training of LSSVM. The program of LSSVM has been developed by using MATLAB.

4. Results and Discussion

The coefficient of correlation (R) is the main criterion to access the performance of LSSVM model. The value of R has been determined by using the following equation:

$$R = \frac{\sum_{i=1}^{n}\left(C_{cai} - \overline{C}_{ca}\right)\left(C_{cpi} - \overline{C}_{p}\right)}{\sqrt{\sum_{i=1}^{n}\left(C_{cai} - \overline{C}_{ca}\right)}\sqrt{\sum_{i=1}^{n}\left(C_{cpi} - \overline{C}_{cp}\right)}} \tag{5}$$

Where C_{cai} and C_{cpi} are the actual and predicted C_c values, respectively, $\overline{C}_{ca}$ and $\overline{C}_{cp}$ are mean of actual and predicted C_c values corresponding to n patterns. The design values of γ and σ have been presented in Table 1. The performance of training and testing dataset for single input variable has been depicted in figures 1 and 2. It is observed from figures 1 and 2 that the performance of LSSVM is quite well for single input variable. Equations (by putting

$K(x_i, x) = \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2\sigma^2}\right\}$, N=180, different σ values for ω_1, ω_n, e_0 , PI and different b values for ω_1, ω_n, e_0 , PI in equation 3) have been developed based on single input variable from the LSSVM model. Table 1 shows the different equations for prediction of C_c based on different single input variables. The values of α have been given in figure 3.

For multiple input variables, the design values of γ and σ have presented in Table 1. The performance of training dataset has been determined by using the design value of γ and σ. Figure 4 illustrates the performance of training dataset. The performance of testing dataset has been also depicted in figure 4. The performance of LSSVM is quite well for determination of C_c based on the multiple input variables. Equations (by putting $K(x_i, x) = \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2\sigma^2}\right\}$, N=180, σ=0.09 and b =0.2773 in equation 3) have been developed based on single input variable from the LSSVM model. Table 1 shows the equation for determination of C_c based on the multiple input variables.

Table 1. Performance of the different LSSVM models.

Input variable	Design value of γ	Design value of σ	Training Performance (R)	Testing Performance (R)	Equation
ω_1	100	0.04	0.752	0.749	$C_c = \sum_{i=1}^{180} \alpha_i \exp\left\{\frac{(x_i - x)(x_i - x)^T}{0.0032}\right\} + 0.6009$
ω_n	50	0.05	0.795	0.718	$C_c = \sum_{i=1}^{180} \alpha_i \exp\left\{\frac{(x_i - x)(x_i - x)^T}{0.005}\right\} + 0.0050$
e_0	50	1	0.689	0.661	$C_c = \sum_{i=1}^{180} \alpha_i \exp\left\{\frac{(x_i - x)(x_i - x)^T}{2}\right\} + 0.5940$
PI	20	0.01	0.770	0.719	$C_c = \sum_{i=1}^{180} \alpha_i \exp\left\{\frac{(x_i - x)(x_i - x)^T}{0.002}\right\} + 0.5940$
PI, ω_1 and e_0	60	0.09	0.998	0.730	$C_c = \sum_{i=1}^{180} \alpha_i \exp\left\{\frac{(x_i - x)(x_i - x)^T}{0.0162}\right\} + 0.2773$

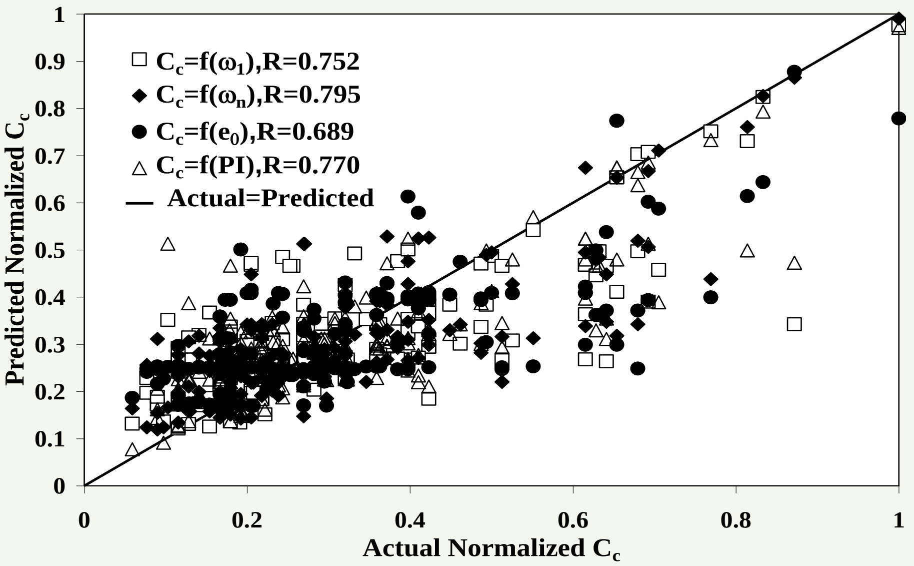

Fig.1: Performance of training dataset for single input variable.

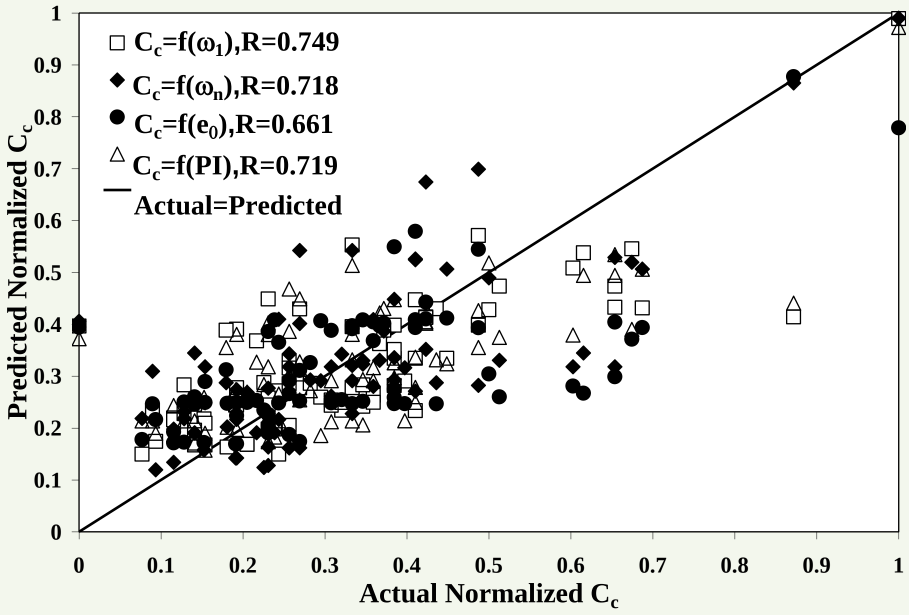

Fig. 2: Performance of testing dataset.

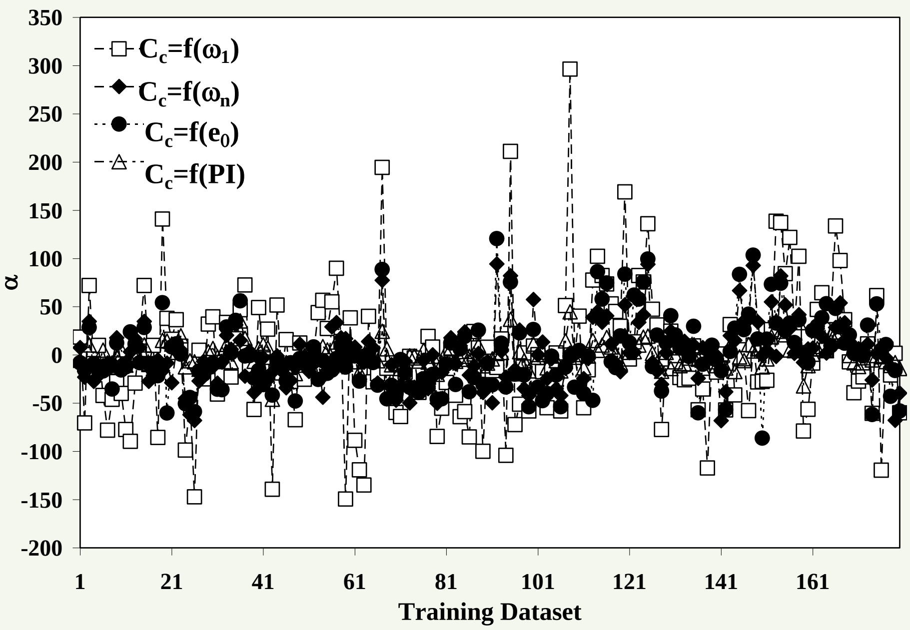

Fig. 3: Values of α for single input variable.

Fig. 4: Performance of LSSVM model for multiple input variables.

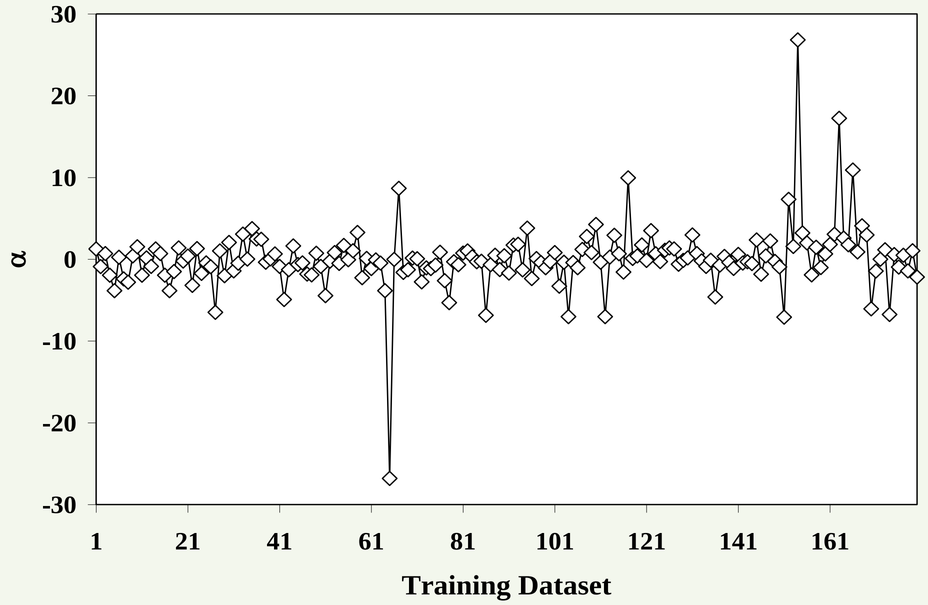

Fig.5: Values of α for multiple input variables.

An comparative study has been carried out between the developed LSSVM models and the regression models developed by Yoon et al.(2004). Comparison has been done for testing dataset. Figure 6 shows the bar chart of R values for the regression and the LSSVM models. It is clear from figure 6 that the performance of the developed LSSVM models is better than the regression models.

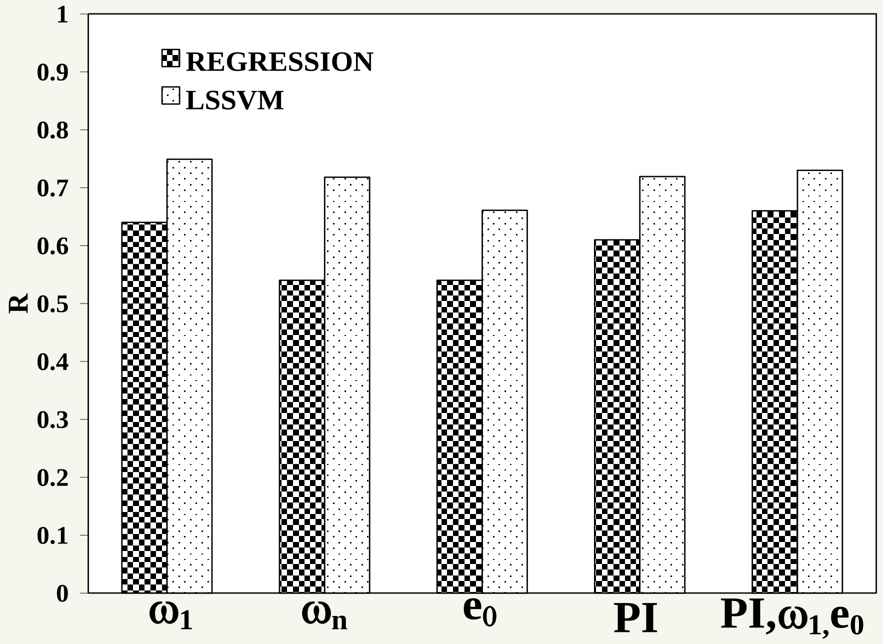

Fig.6 : Bar chart of R for the regression and LSSVM model.

5. Conclusion

This study has presented LSSVM model for determination of C_c of marine clay in east coast of Korea. The LSSVM structure that we have built had given very promising results in predicting C_c of marine clay in east coast of Korea. The developed LSSVM models outperform the regression models. User can use the developed equation for prediction of C_c of marine clay in east coast of Korea. Due to the use of statistical learning theory, LSSVM can avoid several disadvantages such as local optimal solution, low convergence rate, especially poor generalisation when few samples are available, etc. In summary, it can be concluded that the developed LSSVM is a robust model for determination of C_c of marine clay in east coast of Korea.

References

[1] Baylara, A., Hanbay, D. and Batan, M., "Application of least square support vector machines in the prediction of aeration performance of plunging overfall jets from weirs". *Expert Systems with Applications*, Vol.36, No.4, (2009), pp. 8368-8374.

[2] Haykin, S. *Neural networks*, 2nd Ed., Prentice-Hall, Upper Saddle River, N.J., (1999).

[3]Mayne, P.W. "Cam-clay predictions of undrained strength". *Journal of the Geotechnical Engineering Division*, American Society of Civil Engineering, Vol.106, No.11, (1980), pp. 1219–1242,.

[4] Nakase, A., Kamei, T., Kusakabe, O. "Constitute parameters estimated by plasticity index". *Journal of Geotechnical Engineering, American Society of Civil Engineering*, Vol.114, No.GT7, (1988), pp. 844–858.

[5] Samui, P., Sitharam, T.G. "Least-square support vector machine applied to settlement of shallow foundations on cohesionless soils". *International Journal of Numerical and Analytical Method in Geomechanics*, Vol.32, No. 17, (2008) pp. 2033-2043.

[6] Skempton, A.W. "Notes on the compressibility of clays". *Quarterly Journal of the Geological Society of London*, Vol. 100, No.2, (1944), pp. 119–135.

[7] Smola, A., Scholkopf, B. 1998. "On a kernel based method for pattern recognition, regression, approximation and operator inversion". *Algorithmica* 22, (1998), pp. 211-231.

[8] Suykens, J.A.K., Van, G.T., De, B.J., De Moor B., Vandewalle, J. *Least squares support vector machines*, World Scientific, Singapore, (2002).

[9] Terzaghi, K., Peck, R.B. *Soil mechanics in engineering practice*, 2nd ed. John

Wiley & Sons Inc., New York, (1967).

[10] Vapnik, V. The nature of statistical learning theory, Springer-Verlag, New York, (1995).

[11] Vapnik, V. *Statistical learning theory*, John Wiley & Sons, New York, (1998).

[12] Wang, X.D., Ye, M.Y., 2004 “Nonlinear dynamic system identification using least squares support vector machine regression”. *Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai*, (2004),pp. 26-29.

[13]Yin, J.H. “Properties and behaviour of Hong Kong marine deposits with different clay contents”. *Canadian Geotechnical Journal*, Vol. 36, (1999), pp.1085–1095.

[14] Yoon, G.L., Kim, B.T., Jeon, S.S. “Empirical correlations of compression index for marine clay from regression analysis”. *Canadian Geotechnical Journal*, Vol. 41, (2004),pp. 1213–1221.