Automation in Construction 68 (2016) 65–80

Contents lists available at ScienceDirect

Automation in Construction

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

Metaheuristic optimization within machine learning-based classification system for early warnings related to geotechnical problems

Jui-Sheng Chou *, Julian Pratama Putra Thedja

Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 10 October 2015
Received in revised form 22 March 2016
Accepted 26 March 2016
Available online xxxx

Keywords:
Expert system
Early warning
Metaheuristic optimization
Swarm intelligence
Machine learning
Classification
Human–machine interface
Gehazards
Geotechnical engineering

ABSTRACT

This study proposes a novel classification system integrating swarm and metaheuristic intelligence, i.e., a smart firefly algorithm (SFA), with a least squares support vector machine (LSSVM). Benchmark functions were used to validate the optimization performance of the SFA. The experimental results showed that the SFA obtained 100% success rate in searching the optimum for most benchmark functions. The SFA was then integrated with the LSSVM to create a metaheuristic optimized classification model. A graphical user interface was developed for the proposed classification system to assist engineers and researchers in executing advanced machine learning tasks. The system was applied to several geotechnical engineering problems that involved measuring the groutability of sandy silt soil, monitoring seismic hazards in coal mines, predicting postearthquake soil liquefaction, and determining the propensity of slope collapse. The prediction problems in these studies were complex because they were dependent on various physical factors, and such factors exhibited highly nonlinear relations. The analytical results revealed that the metaheuristic optimization within machine learning-based classification system exhibited a groutability prediction accuracy of 95.42%, seismic prediction accuracy of 93.96%, soil liquefaction prediction accuracy of 95.18%, and soil collapse prediction accuracy of 95.45%. Hence, the proposed system is a promising tool to provide decision-makers with timely warnings of geotechnical hazards.

Abbreviations and symbols

Abbreviations

ACO	Ant colony optimization
AI	Artificial intelligence
ANN	Artificial neural network
BCGP	Bayesian classifier for groutability prediction
BRACID	Bottom-up induction of rules and cases for imbalanced data
DE	Differential evolution
DM	Data mining
FA	Firefly algorithm
FEM	Finite element method
GA	Genetic algorithm
GUI	Graphical user interface
LSSVM	Least square support vector machine
MA	Metaheuristic algorithm
NBC	Naïve Bayesian classifier
PSO	Particle swarm optimization
QP	Quadratic programming
RBF	Radial basis function
RBFNN	Radial basis function neural network
SFA	Smart firefly algorithm
SFA–LSSVM	Smart firefly algorithm–least squares support vector machine
SVM	Support vector machine

Symbols

a	Biotic potential in chaotic maps
a_{max}	Horizontal ground surface acceleration
$Accuracy^{Learning\ process}_{Validating-data}$	Prediction accuracy from validation data
b	SVM bias term
C	Regularization parameter of LSSVM
C_u	Uniformity coefficient
C_z	Coefficient of gradation
$D_{10}(BaseSoil)$	Diameter at which 10% of the soil mass passes through the sieve
$D_{15}(BaseSoil)$	Diameter at which 15% of the soil mass passes through the sieve
$d_{85}(CementGrout)$	Diameter at which 85% of the total grout mass passes through the sieve

* Corresponding author. Tel.: +886 2 2737 6321; fax: +886 2 2737 6606.
E-mail addresses: jschou@mail.ntust.edu.tw (J.-S. Chou), m10305812@mail.ntust.edu.tw (J.P.P. Thedja).

http://dx.doi.org/10.1016/j.autcon.2016.03.015
0926-5805/

$D_{90}(CementGrout)$	Diameter at which 90% of the total grout mass passes through the sieve
$D_{95}(CementGrout)$	Diameter at which 95% of the total grout mass passes through the sieve
e	Void ratio
e_i	Relaxation vector in which e_i is the relaxation variable for the ith sample
$f(m)$	SFA–LSSVM objective function
FC	Fines content of total soil mass
fn	False negative
fp	False positive
$\Gamma(z)$	Gamma function
i	ith iteration of the optimization process
i_{max}	Maximum iteration of the optimization process
$I(r)$	Light intensity of firefly algorithm
I_s	Intensity at the source of firefly algorithm
$K(\vec{x}, \vec{x}_i)$	Kernel function
$L(\alpha_i)$	Lagrange multiplier of α_i variable
$L(\vec{w}, b, e, \alpha_i)$	Lagrange multiplier of $\vec{w}, b, e, \alpha_i$ variable
$L(s)$	Lévy distribution
LB	Lower bound
mod	Remainder after division (modulo operation)
M_w	Earthquake movement magnitude
N	Incecik and Ceran soil groutability index
N_1	1st Burwell soil groutability index
N_2	2nd Burwell soil groutability index
N_{all}	Number of total trial
$N_{successful}$	Number of successful trial
q_c	Cone tip resistance
r	Distance between any two fireflies
$rand$	Uniformly distributed random number between 0 and 1
R_f	Sleeve friction ratio
s	Power-law distribution
$sign$	Signum function
S_r	Success rate
tn	True negative
tp	True positive
UB	Upper bound
$\vec{w}$	SVM margin vector
w/c	Water-to-cement ratio of grout
$\vec{x}_+$	SVM positive class vector
$\vec{x}_-$	SVM negative class vector
x_k	Firefly k
x_j	Firefly j
X^o	Default value of data attributes
X^*	SFA optimization result
X^{gb}	Global optima variable of benchmark function
X^{min}	Minimum value of data attributes
X^{max}	Maximum value of data attributes
X^n	Value of data attributes after min-max normalization
y_i	Indicating the class to which the point $\vec{x}$ belongs
Z_n	Chaotic value of nth firefly
α	Firefly random movement coefficient
α_0	Initial firefly random movement coefficient (α)
α_i	SVM alpha dot products
β	Attractiveness coefficient of firefly algorithm
β_0	Minimum value of attractiveness coefficient (β)
β_{chaos}	Chaos attractive parameter β from Gauss map
∂	Differential sign
σ	Sigma parameter for RBF kernel
σ_v	Total stress
σ_v'	Effective stress
γ	Absorption coefficient
θ	Adaptive Inertia weight
τ	An index in Lévy flight
τ_{oct}	Octahedral shear stress (kPa)
$\otimes$	Entrywise multiplications
$\cdot$	Dot products

1. Introduction

Machine learning (ML) techniques have potential for solving numerous real-world problems [1,8,9,11,28,55] and are currently an essential research area. Supervised classification is a prominent machine learning process for extracting information from a data set to create a model that can be used to make predictions. Most ML studies have focused on developing artificial intelligence (AI) algorithms for increasing classification accuracy levels. Soft computing algorithms have also demonstrated superior predictive abilities compared with conventional methods [13]. However, most of such algorithms are not easy to use for simulating complex behaviors in engineering issues. AI-embedded systems can be computer-aided tools to alleviate the mentioned concerns. Accordingly, researchers have studied engineering behaviors periodically and developed expert systems to reflect real-world conditions.

Geohazards are among the complex and heterogeneous engineering problems that deserve further investigation. For example, earthquakes can cause seismic bumps in coal mines, soil liquefaction, and slope collapse. Unpredictable factors and numerous variables must be evaluated when deriving solutions for these complex geotechnical problems. Moreover, natural disasters can cause various consequences and losses. Numerous lives can be saved by implementing an early warning system that can provide alerts a few seconds to a minute before the occurrence of disasters. When a system broadcasts an early warning of earthquake events, the evacuation efficiency at hazardous locations can be enhanced. Additionally, an early warning system implemented at the planning stages can fully minimize the project lifecycle loss of life, material, and effort. Thus, proactive planning or automated actions can be taken to mitigate impending impacts of secondary hazards.

Although advanced AI techniques have numerous advantages in early prediction, geotechnical engineers still face difficulties in understanding and implementing them in practical field applications. Therefore, developing an expert system with a user-friendly interface is imperative to facilitate decision-making process. The objective of the current study was to fill the knowledge gap by designing a computer-aided system in which users can efficiently implement the advanced AI algorithms with their domain specialty. Thus, a human-machine interface, called nature-inspired metaheuristic classification system, was developed in this study for resolving general geotechnical problems. The performance of the proposed system was then validated by applying it to geotechnical case studies and compared with empirical methods and previous works.

The support vector machine (SVM) algorithm is one of the prominent AI algorithms. Several studies have applied SVM algorithms to solve geotechnical problems, including forecasting building settlements after soil liquefaction [54], predicting the California bearing ratio of a stabilized expansive soil [43], and evaluating the stability of rock slopes [61]. The least squares SVM (LSSVM) algorithm, which is a modification of the SVM algorithm, reduces computational complexity by applying a least squares formulation [51]. However, the parameters of either SVM or LSSVM algorithm must be fine-tuned to enhance prediction accuracy. The performance of the LSSVM algorithm depends on the selected penalty parameter and kernel function parameter [37], which are called LSSVM hyperparameters.

Obtaining a set of LSSVM hyperparameters that is ideal for every scenario is impractical. Therefore, an optimization algorithm must be integrated with the LSSVM algorithm for fine-tuning the hyperparameters to prevent the existence of over-fitting and local minima problems, improving the prediction accuracy. Thus, this study integrated the LSSVM with a swarm optimization algorithm for effectively predicting the occurrence of inherently complex geotechnical problems. Researchers