Technical Note

Predicting Compressive Strength of High-Performance Concrete Using Metaheuristic-Optimized Least Squares Support Vector Regression

Anh-Duc Pham, Ph.D.[1]; Nhat-Duc Hoang, Ph.D.[2]; and Quang-Trung Nguyen[3]

Abstract: This research establishes a novel model for predicting high-performance concrete (HPC) compressive strength, which hybridizes the firefly algorithm (FA) and the least squares support vector regression (LS-SVR). The LS-SVR is utilized to discover the functional relationship between the compressive strength and HPC components. To achieve the most desirable prediction model that features both modeling accuracy and generalization capability, the FA is employed to optimize the LS-SVR. To construct and verify the proposed model, this study has collected a database consisting of 239 HPC strength tests from an infrastructure development project in central Vietnam. Experimental results have demonstrated that the new model is a promising alternative to predict HPC strength. **DOI: 10.1061/(ASCE)CP.1943-5487.0000506.** *© 2015 American Society of Civil Engineers.*

Author keywords: Artificial intelligence; Compressive strength; Firefly algorithm; High-performance concrete; Machine learning.

Introduction

The compressive strength is often considered as the most important property of high-performance concrete (HPC); other concrete properties such as elastic modulus and water tightness appear to have direct relationships with the compressive strength. Hence, the compressive strength is commonly used as the main criterion in defining the required quality of concrete (Ahmadi-Nedushan 2012), and an accurate estimation of the strength before the placement is a practical need of construction engineers.

Due to the importance of the research topic, modeling HPC compressive strength using artificial intelligence (AI) has been a very active research area (Castelli et al. 2013; Cheng et al. 2014; Chou et al. 2011; Erdal et al. 2013; Kasperkiewicz et al. 1995; Yeh 1998). Based on previous studies, AI techniques have proved their superior capability over traditional modeling methods. This research extends the body of knowledge by proposing a novel approach for predicting HPC compressive strength based on the LS-SVR. The LS-SVR is an advanced AI method that is suitable for nonlinear modeling (Suykens et al. 2002). Although the superiority of the approach has been strongly demonstrated in various recent applications (Cheng and Hoang 2014), none of the previous research has investigated the capability of the LS-SVR for predicting HPC strength. Therefore, the current study is an attempt to fill this gap.

To optimize the LS-SVR learning process, this work employs the firefly algorithm (FA) (Yang 2008), which is a fast and effective metaheuristic. Additionally, to establish the model, this research has collected HPC strength tests from an infrastructure development project. The rest of the article is organized as follows: the first section describes the research method., the proposed prediction model is addressed in the next section, the third section reports the experimental results, and the conclusion of this study is stated in the final section.

Research Method

Collected Data Set of HPC Specimens

This research recorded 239 testing results of HPC samples during the progress of the Nga Ba Hue transport intersection project, in Danang City, Vietnam (http://ngabahue.com.vn). All tests were performed on 15-cm cylindrical specimens of HPC prepared in conformity with standard procedures. The amount of cement (kg/m^3), sand (kg/m^3), small coarse aggregate (kg/m^3), medium coarse aggregate (kg/m^3), water (L/m^3), and superplasticizer (L/m^3) are mixture components used to characterize a concrete sample.

Cement used in the project was produced by four companies (Song Gianh, Hai Van, Nghi Son, and Kim Dinh) in Vietnam. There are four suppliers of aggregates (Da Son, Hoc Khe, Hoa Nhon, and Ha Nha) and three suppliers of admixtures (Mapei Vietnam, Sika Vietnam, and BASF Vietnam). All concrete for the project was supplied by Phuoc Yen Company in Danang City, Vietnam. The concrete specimens were cured in moist air, and the curing process is in compliance with the Vietnamese standard (TCVN 3105-1993) for sampling, making, and curing of concrete test specimens. The temperature and humidity in the laboratory during the concrete curing process are $27°C \pm 1$ and 90–95%, respectively. Furthermore, the concrete age after pouring is also recorded and used as an influencing factor, which affects the concrete strength. Herein, the compressive strength HPC samples are measured at different ages: 3, 7, and 28 days. Moreover, all the concrete samples were tested by the CTES construction company in Danang City, Vietnam.

[1]Lecturer, Faculty of Project Management, Univ. of Danang—Univ. of Science and Technology, 54 Nguyen Luong Bang, Danang 550000, Vietnam. E-mail: paduc@dut.udn.vn

[2]Lecturer, Institute of Research and Development, Faculty of Civil Engineering, Duy Tan Univ., P809—K7/25 Quang Trung, Danang 550000, Vietnam (corresponding author). E-mail: hoangnhatduc@dtu.edu.vn

[3]Lecturer, Faculty of Project Management, Univ. of Danang—Univ. of Science and Technology, 54 Nguyen Luong Bang, Danang 550000, Vietnam. E-mail: nqtrung@dut.udn.vn

Note. This manuscript was submitted on November 26, 2014; approved on April 10, 2015; published online on May 26, 2015. Discussion period open until October 26, 2015; separate discussions must be submitted for individual papers. This technical note is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801/06015002(4)/$25.00.

Table 1. Concrete Components and Statistical Descriptions

Factors	Notation	Minimum	Mean	Standard deviation	Maximum
Cement (kg/m^3)	X_1	350.0	447.4	25.0	498.0
Fine aggregate (kg/m^3)	X_2	666.0	728.9	37.8	879.0
Small coarse aggregate (kg/m^3)	X_3	0.0	347.2	55.9	424.0
Medium coarse aggregate (kg/m^3)	X_4	626.0	721.3	61.4	1,060.0
Water (L/m^3)	X_5	134.0	178.8	20.4	207.0
Superplasticizer (L/m^3)	X_6	3.5	5.1	0.6	7.0
Concrete age (day)	X_7	3.0	15.1	10.9	28.0
Compressive strength (MPa)	Y	23.6	42.5	13.5	85.2

Statistical descriptions of all concrete samples are provided in Table 1. The data set is illustrated in Table 2. It is noted that the small coarse aggregate has the diameter ranging from 5 to 10 mm; the medium coarse aggregate has the diameter varying between 10 and 20 mm. Furthermore, the water-to-cement ratio in the data set ranges from 0.27 to 0.46; such low water-to-cement ratios necessitate the use of superplasticizer to improve the concrete workability. The minimum and maximum compressive strength of concrete samples are 23.6 and 85.2 MPa, respectively.

Least Squares Support Vector Regression

The learning objective of the least squares support vector regression (LS-SVR) can be transformed into the following optimization problem (Suykens et al. 2002):

$$\text{Minimize } J_p(w,e) = \frac{1}{2}w^T w + \gamma \frac{1}{2}\sum_{k=1}^{N} e_k^2 \quad \text{subjected to}$$
$$y_k = w^T \varphi(x_k) + b + e_k, \quad k = 1, \ldots, N \tag{1}$$

where$e_k \in R$ = error variables; and $\gamma > 0$ denotes a regularization constant.

To solve the optimization problem, the Lagrangian can be formulated and using Karush–Kuhn–Tucker (KKT) conditions for optimality, the resulting model is expressed as (Suykens et al. 2002)

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x_k, x_l) + b \tag{2}$$

where α_k and b = the solution to the linear system; k and N = the index and the total number of data points in the training set; x_k and x_l = an input pattern in the training and testing set; and $K(.)$ = the kernel function that maps the input data from the feature space into the high-dimensional space. The radial basis function (RBF) kernel is often utilized

$$K(x_k, x_l) = \exp\left(\frac{\|x_k - x_l\|^2}{2\sigma^2}\right) \tag{3}$$

where σ = the RBF kernel function parameter.

Firefly Algorithm

The firefuly algorithm (FA) is an optimization algorithm inspired by flashing behaviors of fireflies (Yang 2008). The flash patterns, produced by a bioluminescence process, are used by the fireflies for attracting potential preys and mating partners. The algorithm implementation is based on a physical formula of light intensity that decreases with the increase of the distance. The FA algorithm can be illustrated in Fig. 1. For the detailed description of the FA, the readers are guided to the previous works of Yang (Yang 2008, 2014).

Proposed HPC Compressive Strength Prediction

This section describes the proposed HPC compressive strength prediction model, which is named as FA-LSVR. The flowchart of the model is provided in Fig. 2.

1. Input data: the whole data set is randomly separated into the training and the testing sets;
2. Tuning parameter initialization: the aforementioned tuning parameters of the model are randomly generated within the range of lower boundary (10^{-5}) and upper boundary (10^5);
3. Model training: the LS-SVR is deployed to generalize the nonlinear function that governs the input-output mapping between the concrete components and the compressive strength;
4. Fitness evaluation: to measure both accuracy and generalization of the model, the following objective function is used:
$$F_{\text{fitness}} = \frac{\Sigma_{k=1}^{S} E_{tr}^k}{S} + \frac{\Sigma_{k=1}^{S} E_{va}^k}{S} \tag{4}$$
where E_{tr}^k and E_{va}^k = the training and validating errors, respectively, for kth run; and $S = 5$ = the total number of data folds. The training and validating errors are root mean squared error (RMSE);
5. Firefly algorithm searching: the FA automatically explores the search space to evaluate the various combinations of the hyperparameters (γ and σ); and
6. Optimized FA-LSVR: when the searching process stops, the optimized prediction model has been identified.

Table 2. Data set of HPC

HPC specimen	X_1 (kg/m^3)	X_2 (kg/m^3)	X_3 (kg/m^3)	X_4 (kg/m^3)	X_5 (L/m^3)	X_6 (L/m^3)	X_7 (Day)	Y (MPa)
1	400	797	342	798	141	4.92	3	23.60
2	460	732	324	756	168	5.52	3	25.10
3	440	691	318	742	202	4.84	28	28.78
4	449	681	317	739	200	4.94	7	29.54
5	350	877	318	742	158	3.50	7	30.90
…	…	…	…	…	…	…	…	…
235	460	751	315	735	172	5.52	7	54.90
236	475	735	321	749	166	5.46	3	55.10
237	498	762	320	746	149	6.23	28	77.80
238	498	754	321	749	149	6.97	28	84.30
239	498	767	324	756	140	6.97	28	85.20

Begin algorithm
Define the objective function $f(s)$, where $s=(s_1, s_2,...,s_d)$
Generate an initial swarm of fireflies
Formulate the light intensity I
Define the absorption coefficient γ
While (t<Maximum Generation)
For i = 1 to n (all n fireflies)
For j=1 to n (all n fireflies)
If (I_j>I_i) the i^{th} firefly moves towards the j^{th} firefly
End if
Evaluate new solutions and update the light intensity;
End for j
End for i
Rank the fireflies and identify the current best solution
End while;
End algorithm

Fig. 1. Firefly algorithm

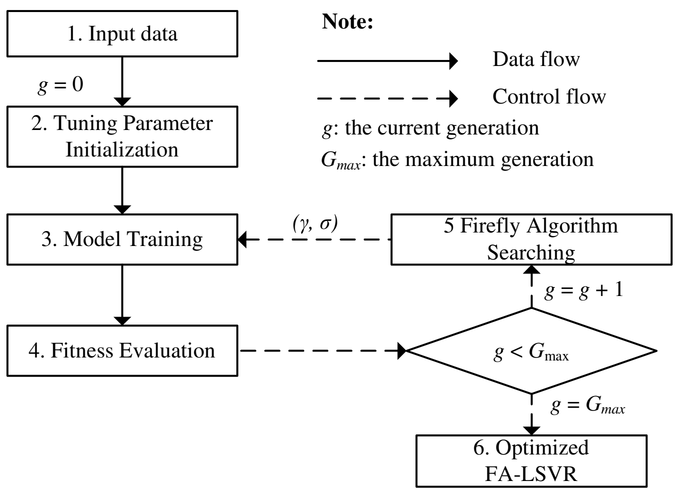

Fig. 2. Firefly algorithm optimized least squares support vector regression

Model Application

Experiment Setting

To verify the performance of the proposed FA-LSVR, the artificial neural network (ANN) and the SVM are used as benchmark methods. Furthermore, to measure model performance, this research employs RMSE, mean absolute percentage error (MAPE), and coefficient of determination (R^2). On the basis of experiments, the ANN configuration is described as follows: the number of hidden layers = 1; the learning rate = 0.001; the number of neurons = 7. The Levenberg-Marquardt algorithm is employed to train the ANN (Hagan and Menhaj 1994). Moreover, the hyperparameters of the SVM are fine-tuned by the FA.

Experimental Results

In the first experiment, the data set is randomly divided into two sets: training set (90%) and testing set (10%). In detail, the training and testing sets consist of 215 and 24 samples, respectively. Fig. 3 shows the actual concrete slumps against their predicted values; the experimental result indicates that the proposed FA-LSVR has obtained a good fit to a straight line. Results of the first experiment are summarized in Table 3. Observably, the FA-LSVR has achieved the highest R^2 (0.89) and the lowest MAPE (9.38%) and RMSE (4.82). The SVM is the second best approach, followed by the ANN.

In the next experiment, a tenfold cross validation process is performed. Because all of the subsamples are mutually exclusive, this process can better evaluate the proposed FA-LSVR and other benchmarking methods. Table 4 summarizes results of the second experiment. The testing RMSE and MAPE of the proposed method are 4.86% and 9.81%, respectively. These results are significantly better than the SVM, which is the second-best model. In terms of MAPE, the FA-LSVR has obtained roughly 18% and 27% reductions compared with the SVM and ANN models, respectively. The proposed approach also yields the highest R^2 (0.87) when predicting the testing data.

Discussion

In the current research, the main objective is to establish a LS-SVR-based machine learning model that can well generalize the mapping between HPC strength and its influencing factors. According to the experimental result, the proposed FA-LSVR can successfully predict HPC strength since it has achieved a comparatively low MAPE. This implies that on average, the absolute deviation between the predicted and actual concrete strength is less than 10%, which is very desirable because the problem at hand is known to be highly complex and inherently uncertain. The complexity of the HPC

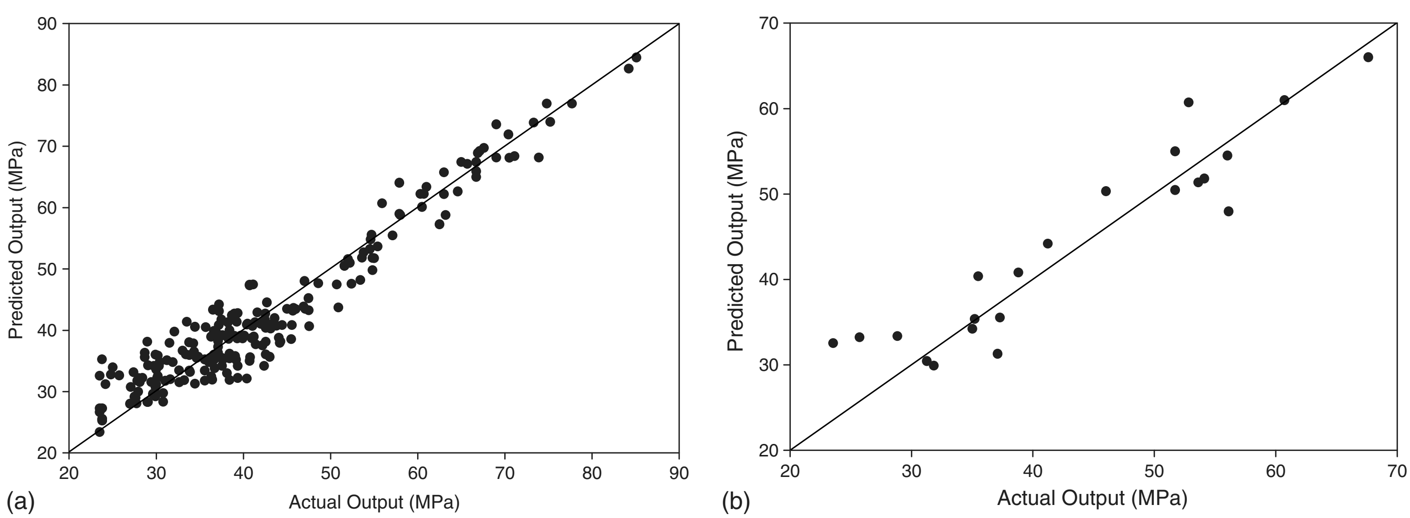

Fig. 3. Actual versus predicted HPC strengths: (a) training ($R^2 = 0.92$); (b) testing ($R^2 = 0.89$)

Table 3. Result Comparison

Model	Performance	ANN	SVM	FA-LSVR
Train	RMSE	5.63	2.96	3.79
	MAPE	12.00	4.49	8.15
	R^2	0.82	0.95	0.92
Test	RMSE	6.39	5.89	4.82
	MAPE	12.39	11.37	9.38
	R^2	0.81	0.83	0.89

Table 4. Average Results of the Tenfold Cross-Validation Process

Model	Performance	ANN	SVM	FA-LSVR
Train	RMSE	5.32	4.73	3.34
	MAPE	10.75	8.20	6.76
	R^2	0.83	0.86	0.93
Test	RMSE	6.74	6.07	4.86
	MAPE	13.41	12.02	9.81
	R^2	0.76	0.79	0.87

strength prediction is due to the nonlinear and multivariate nature of the learning problem. Furthermore, the wide ranges of suppliers of materials (cements, aggregates, and superplasticizers) as well as variations in concrete mixing procedures and levels of quality control are the explanations for the uncertainty of the results.

Additionally, in regression analysis, the case in which the actual and predicted outcomes are exactly the same hardly exists. Thus, the predicted strength is either greater than the actual strength (Case 1) or smaller than the actual strength (Case 2). Moreover, since the objective is to generalize the input-output mapping with the assumption that the data-collection process is unbiased, the rates of Case 1 and Case 2 are expected to be 50%.

Indeed, the tenfold cross-validation process reveals that 46.5% of the prediction results belong to Case 1. This is understandable because the regression machine is designed to model the nonlinear mapping as accurately and generally as possible. However, in the problem of HPC strength prediction, the outcomes of Case 1 are more undesirable than that of Case 2. Thus, the users should be aware of the situation of Case 1, in which the predicted strength is greater than the actual strength, when they estimate strengths of HPC mixtures using the proposed FA-LSVR.

Conclusions

This research has proposed and verified the FA-LSVR for estimating HPC compressive strength. From the experimental results, the proposed method has achieved the most desirable performance with low prediction errors. The FA-LSVR is the approach that is deemed best suited for the problem of interest. From the perspective of construction engineers, the proposed FA-LSVR can be a useful tool to identify potentially unqualified HPC mixtures in a timely manner. The future directions of the current research include investigating the effect of the concrete temperature on the HPC compressive strength and developing novel machine learning techniques to improve the prediction accuracy as well as to reduce the undesirable situation in which the estimated strength is greater than the actual strength.

References

Ahmadi-Nedushan, B. (2012). "An optimized instance based learning algorithm for estimation of compressive strength of concrete." *Eng. Appl. Artif. Intell.*, 25(5), 1073–1081.

Castelli, M., Vanneschi, L., and Silva, S. (2013). "Prediction of high performance concrete strength using genetic programming with geometric semantic genetic operators." *Expert. Syst. Appl.*, 40(17), 6856–6862.

Cheng, M., and Hoang, N. (2014). "Risk score inference for bridge maintenance project using evolutionary fuzzy least squares support vector machine." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000275, 04014003.

Cheng, M.-Y., Firdausi, P. M., and Prayogo, D. (2014). "High-performance concrete compressive strength prediction using genetic weighted pyramid operation tree (GWPOT)." *Eng. Appl. Artif. Intell.*, 29, 104–113.

Chou, J.-S., Chiu, C.-K., Farfoura, M., and Al-Taharwa, I. (2011). "Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000088, 242–253.

Erdal, H. I., Karakurt, O., and Namli, E. (2013). "High performance concrete compressive strength forecasting using ensemble models based on discrete wavelet transform." *Eng. Appl. Artif. Intell.*, 26(4), 1246–1254.

Hagan, M. T., and Menhaj, M. B. (1994). "Training feedforward networks with the Marquardt algorithm." *IEEE Trans. Neural Networks*, 5(6), 989–993.

Kasperkiewicz, J., Racz, J., and Dubrawski, A. (1995). "HPC strength prediction using artificial neural network." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)0887-3801(1995)9:4(279), 279–284.

Suykens, J., Gestel, J. V., Brabanter, J. D., Moor, B. D., and Vandewalle, J. (2002). *Least square support vector machines*, World Scientific Publishing, Singapore.

Yang, X.-S. (2008). *Firefly algorithm*, Luniver Press, Bristol, U.K.

Yang, X.-S. (2014). *Nature-inspired optimization algorithms*, Elsevier, Amsterdam, Netherlands.

Yeh, I.-C. (1998). "Modeling of strength of high-performance concrete using artificial neural networks." *Cem. Concr. Res.*, 28(12), 1797–1808.