Predicting Maximum Crest Settlement of Concrete Face Rockfill Dams Using a New Ensemble Learning Model

Lifeng Wen\textsuperscript{1*}, Haiyang Zhang\textsuperscript{1} and Yanlong Li\textsuperscript{1}

\textsuperscript{1}State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China, Xi’an University of Technology, Xi’an, Shaanxi, 710048, P.R. China

*Corresponding author’s e-mail: wenlifeng@xaut.edu.cn

Abstract: Deformation assessment and control are essential issues in the construction of concrete face rockfill dams (CFRDs). The design and construction of CFRDs require deformation behavior that can be estimated rapidly to support engineering optimization and safety assessment. Based on 87 case histories of in-service CFRDs, a new ensemble learning model has been developed to predict maximum crest settlement (CS) of CFRDs. The model is based on the support vector machine regression algorithm (SVR) combined with multiple variables, and then the foundation model is integrated by the weighted average integration method. It is demonstrated here that the new ensemble learning model weakens the nonlinear characteristics of case data, makes up for the instability of single regression algorithm, improves the generalization ability, and provides a new idea for predicting the CS of CFRDs.

1. Introduction

With their low costs, strong adaptability to geological conditions, and the ability to be constructed from local materials, concrete face rockfill dams (CFRDs) have become one of the most competitive dam types [1]. The construction of CFRDs requires the maximum crest settlement (CS) to be estimated rapidly to support engineering optimization and safety assessment.

The empirical equations generally relate the CS of a CFRD to only one or two influencing factors. The regression model’s functional form was relatively simple, and thus, it could not always accurately fit the discrete deformation data. In order to overcome the shortcomings of empirical prediction methods, intelligent algorithms are used continuously in the prediction of the CS. Li JF et al. [2] studied the neural network model’s application in the settlement deformation of CFRDs during the construction period; Kim and Kim [3] established an intelligent prediction model for relative crest settlement based on the 30 CFRDs. The neural network model for predicting the CS in CFRDs is easy to fall into a local minimum, and the convergence speed of the learning process is slow. Support vector machine regression (SVR) has been extended to solve the nonlinear regression estimation problem. Besides, there are many factors affecting the CS of CFRDs, and the database data features are complex. Therefore, it is necessary to study further the prediction model of the CS of CFRDs.

This paper aims to develop robust prediction models for predicting the CS of CFRDs based on numerous case histories. Compared with a single SVR, the ensemble learning algorithm has higher accuracy and more significant generalization performance. Based on these considerations, the correlation analysis method is used to determine the correlation coefficient between the CS and five dam-construction-related influencing factors (dam height, foundation behavior, valley shape factor, void ratio, and operation time). Three SVR (SVR-3—SVR-5) prediction models were established by selecting the influencing factors in order of correlation coefficients. The weighted average (WA)
Integration method was used to integrate several SVR models' prediction results with good prediction accuracy. The integration results show that the WA-SVR model has good prediction accuracy and generalization ability.

2. Behavioral characteristics of CFRDs

The deformation of CFRDs occurs during construction, water storage, and operation. As the dam is continuously filled, most of the settlement deformation occurs during the construction stage under self-weight action. After the reservoir filling, the fragmentation, rearrangement, stress release, adjustment, and transfer of rockfill causes creep deformation of the dam [4]. Maximum crest settlement ($CS$) is the essential deformation metrics typically used to evaluate the deformation behaviors of CFRDs. The typical CFRD deformation patterns of four CFRDs are shown in Fig. 1.

The $CS$ in this study refers to the maximum accumulated settlement at the crest of the dam body at the measurement time. Numerous measured results show that the crest settlement mainly occurs in the middle of the dam crest and gradually decreases toward both sides. The $CS$ is mainly composed of time-dependent deformation and deformation caused by reservoir filling. Impoundment accelerates the $CS$ process and causes a significant increase in the crest settlement.

3. Methodology

Based on the SVR principle, the integrated algorithm for the prediction of the $CS$ of CFRDs is based on several SVRs [6]. The specific steps are summarized as follows:

1. Determining model input and output. According to the engineering experience, the main factor $X = \{H, F, e, SF, T\}$ affecting the $CS$ of CFRDs is selected as the input vector. $H$ is dam height, $F$ is foundation condition, $e$ is the void ratio, $SF$ is valley shape, and $T$ is operation time.

2. Calculating the correlation coefficient of influencing factors. The SPSS is used to calculate the correlation coefficients between the $CS$ and $H$, $F$, $e$, $SF$ and $T$, and arranges according to the correlation coefficient.

3. Constructing the prediction model of each SVR. Based on MATLAB, the first three columns of control variables with the largest correlation coefficient are selected as the input variables of the SVR model, and the SVR-3 prediction model is established; the first four columns of control variables with the largest correlation coefficient are selected as the input variables of the SVR model, and the SVR-4 prediction model is constructed. By analogy, the SVR-5 prediction model is built.

4. Training and debugging SVR-3—SVR-5 prediction model. The training samples are used to train and debug the SVR-3—SVR-5 models to determine each model's relevant parameters. The accuracy and generalization ability of the SVR-3—SVR-5 prediction models are tested with the test samples.
5. WA-SVR integration. The average relative error and maximum relative error are used to evaluate the prediction accuracy and generalization ability of SVR-3—SVR-5 models [7]. Different weights are given according to the SVR model's prediction effect. The weighted average value is then calculated, and the weighted average value is taken as the prediction value of the integrated model. The prediction results of several optimal models are selected for WA integration according to the following formula:

\[
\bar{x} = \sum_{i=1}^{k} \omega_i \bar{x}_i
\]

Where: \( \omega_i \) is the weight of the \( i \) model. \( \sum_{i=1}^{k} \omega_i = 1 \) \( \omega_i \geq 0 \) The \( \omega_i \) is determined according to the average relative error of the predicted value of the SVR model. The model with a smaller absolute value of prediction relative error is given greater weight. The calculation formula is as follows:

\[
\omega_i = \frac{1/|e_i|}{\sum_{i=1}^{k} 1/|e_i|}
\]

Where: \( e_i \) is the absolute value of the relative error predicted by the \( i \) model.

6. The prediction results of the WA-SVR model are analyzed. If the predicted value of the WA-SVR model fails to meet the expected accuracy and generalization ability requirements, return to step 4 to debug and verify each SVR individual model until the predicted value of the WA-SVR model meets the expected accuracy requirements.

4. The new ensemble learning model for the CS of CFRDs

4.1. Data collection and preprocessing

This paper is based on the database of measured deformation behavior of CFRDs in reference [5]. In the SVR models developed here, 67% of the available data are used for training, and 33% are reserved for testing the prediction of the CS in CFRDs. At the initial stage of constructing the prediction models, to simplify the design procedure, the prepared database is normalized to within the range of 0–1. The correlations obtained using the SPSS program between the CS and \( H, F, e, SF \), and \( T \) were 0.561, 0.142, 0.173, 0.185 and 0.137, respectively.

4.2. Determination structure for each SVR model

According to the above method, the input variables with different dimensions are selected from the influencing factors according to the order of correlation coefficient, and the SVR-3—SVR-5 models are established. The measured data of 87 CFRDs are fitted and predicted.

After referring to many references, the RBF is selected as the kernel function of SVR. The search space of penalty factor \( C \) and kernel function parameter \( g \) is set as \( 2^{-10} - 2^{10} \), \( K \) is 2-5, the step size of \( C \) and \( g \) is 0.1-0.5, and the insensitive coefficient \( e \) is 0.001-0.1. The \( C \) and \( g \) in the model are determined by the cross-validation method (other parameters adopt system default values).

The best parameter of SVR-3—SVR-5 models are shown in Table 1, and the fitting and prediction results are shown in Fig. 2. It can be seen from Fig. 2 that the prediction accuracy and generalization ability of the SVR model have an apparent improvement trend with the increase of input variable dimension. The mean square error of SVR-3—SVR-5 models for the prediction of test samples is less than 0.1, which has good prediction accuracy and generalization ability. Due to the low dimension of the input vector of the SVR-3 model, the fitting and prediction results of the SVR-3 model are not ideal.

Based on the consideration of prediction accuracy and generalization ability, SVR-3—SVR-5 models with better prediction effects are selected as the individual models of WA-SVR model integration.
Table 1. Optimal parameters of SVR-3—SVR-5 models

| Model | SVR-3 | SVR-4 | SVR-5 |
|-------|-------|-------|-------|
| $C$   | 0.2679| 6.0629| 0.6156|
| $\gamma$ | 0.2176| 0.3536| 2.6390|
| $\varepsilon$ | 0.01 | 0.001 | 0.001 |

![Diagram of prediction effect of SVR-3 model](image1)

(a) Prediction effect of SVR-3 model

![Diagram of prediction effect of SVR-4 model](image2)

(b) Prediction effect of SVR-4 model

![Diagram of prediction effect of SVR-5 model](image3)

(c) Prediction effect of SVR-5 model

Fig. 2. Prediction effect of SVR-3—SVR-5 models
4.3. Weight determination and prediction result analysis of WA-SVR model

Using the mean square error (MSE) of test samples from the SVR-3—SVR-5 models, the weights were determined, and the results are 0.2655, 0.3259, and 0.4086, respectively. The above SVR-3—SVR-5 models with better prediction effects are integrated by the WA integration method, and the WA-SVR model is constructed to predict the database detection samples. The prediction results are compared with SVR-4, SVR-5 individual models. The results are shown in Table 2.

Table 2. Prediction results and comparison of the CS of CFRDs based on WA-SVR model

| No. | Dam   | Crest set. (m)C_τ/10^-4 | WA-SVR     | SVR-4     | SVR-5     |
|-----|-------|-------------------------|------------|-----------|-----------|
|     |       |                         | C_τ/10^-4  | MAE       | C_τ/10^-4 | MAE       | C_τ/10^-4 | MAE       |
| 1   | Sugaroaf | 0.29                    | 0.27       | 0.02      | 0.21      | 0.08      | 0.35      | 0.06      |
| 2   | Fozdo  | 0.08                    | 0.09       | 0.01      | 0.02      | 0.06      | 0.15      | 0.07      |
| 3   | Mangrov| 0.13                    | 0.17       | 0.04      | 0.18      | 0.05      | 0.21      | 0.08      |
| 4   | Alto   | 0.09                    | 0.06       | 0.03      | 0.21      | 0.12      | 0.2       | 0.11      |
| 5   | White  | 0.32                    | 0.36       | 0.04      | 0.33      | 0.01      | 0.27      | 0.05      |
| 6   | Reece  | 0.15                    | 0.18       | 0.03      | 0.25      | 0.1       | 0.11      | 0.04      |

5. Conclusions

Based on the principle of SVR and NNE, this paper constructs the SVR prediction of the CS model with different dimensions by way of multivariable optimal combination. It uses the WA integration method to integrate the optimized foundation model. The results show that the WA-SVR model has high prediction accuracy and generalization ability. This paper can provide a reference for related prediction research in the following two aspects:

First, based on SVR's principle and algorithm, a model and method integrated by multiple SVR individual networks are proposed by using different input dimensions. Second, in practical application, the critical factor determining the prediction accuracy and generalization ability of the model is the complexity of the problem itself. It is difficult to a one-sided view of the first mock exam or the algorithm. Only try different models or algorithms, and debug them repeatedly to achieve the desired prediction results.

Acknowledgments

The research described in this paper was funded by the National Natural Science Foundation of China (Grant No. 51909215, 52039008, 51979224, and 51722907), the Natural Science Basic Research Program of Shaanxi (Program No. 2020JQ-641), the Scientific Research Program Funded by the Shaanxi Provincial Education Department (Program No. 19JS047), the Young Talent fund of University Association for Science and Technology in Shaanxi, China (Program No. 20200417), and China Postdoctoral Science Foundation (Program No. 2020M683527).

References

[1] Sherard JL, Cooke JB. (1987) Concrete-face rockfill dam: I. assessment [J]. Journal of Geotechnical Engineering;113(10):1096-1112.
[2] LI JF, YANG QQ, XU WY. (2007) Application of neural network model to prediction of settlement deformation of rockfill body of CFRD during construction period [J]. Journal of Hohai University. Natural Sciences. 1000-1980 35:5<563:SJWLMX>2.0.TX;2-7.
[3] Kim YS, Kim BT. (2008) Prediction of relative crest settlement of concrete-faced rockfill dams analyzed using an artificial neural network model [J]. Computers and Geotechnics.;35(3):313-322.
[4] ICOLD, (2014). Concrete face rock fill dams concepts for design and construction [C]. Committee on materials for fill dams November.
[5] Wen LF; Chai JR; Xu ZG; Qin Y; LiYL. (2018) A statistical review on the behavior of concrete
face rockfill dams based on case histories [J]. Géotechnique., Online, DOI: https://doi.org/10.1680/jgeot.17.p.095.

[6] Chen X, Yang J, Ye Q, et al. (2011) Recursive projection twin support vector machine via within-class variance minimization[J]. Pattern Recognition, 44(10-11):2643-2655.

[7] Tan X, Wang J, Jin S, et al. (2015) GA-SVR and Pseudo-position-aided GPS/INS Integration during GPS Outage[J]. Journal of Navigation, 68(4):1-19.