Evaluation of Compaction Quality Based on SVR with CFA: Case Study on Compaction Quality of Earth-Rock Dam

Jiajun Wang¹; Denghua Zhong²; Binping Wu³; and Mengnan Shi⁴

Abstract: The compaction quality of earth-rock dam materials is a major concern in the evaluation of earth-rock dams. Current compaction quality assessment methods, such as graphical reports or simple prediction models, are imprecise and can cause unobserved quality assessment defects. These methods do not comprehensively consider factors that affect the compaction quality because they do not integrate heterogeneous construction data sets collected by different data acquisition systems. In this research, a method of assessing compaction quality on the basis of support vector regression (SVR), the chaos-based firefly algorithm, is presented. The assessment method has three stages. In the first stage, a chaotic firefly algorithm (CFA) is proposed to optimize the SVR hyperparameters. In the second stage, a multi-source heterogeneous data integration subsystem based on the compaction monitoring system is designed, in which compaction monitoring data, material source statistical data, and detected data from test pits are integrated. Finally, the optimized SVR is used to evaluate the compaction quality of the storehouse surface. The significance of the proposed method is threefold: first, it integrates both chaos theory and the firefly algorithm to optimize the SVR hyperparameters; second, it integrates heterogeneous construction data, allowing comprehensive consideration of factors that affect the compaction quality; and third, it has high prediction accuracy because it implements structural risk minimization. Compared with current models based on empirical risk minimization, the proposed method performs the best according to several error measures. DOI: 10.1061/(ASCE)CP.1943-5487.0000742. © 2018 American Society of Civil Engineers.

Author keywords: Evaluation of compaction quality; Earth-rock dam; Data integration; Support vector regression; Chaos-based firefly algorithm; Real-time compaction-monitoring system.

Introduction and Background

Compactness is an important indicator in evaluating the construction quality of earth-rock dams. Currently, two evaluation methods are used for the compaction quality of earth-rock dams, namely postevaluation and real-time evaluation. Postevaluation, including spot testing method and nuclear density gauge method (IWHR 2014), refers to compaction evaluation through technical means after the compaction is performed by rolling machines. However, postevaluation may incur additional compactness during the inspection process or leave radioactive materials at the job site. Moreover, because the evaluation occurs after the project, it may influence the construction progress. The real-time evaluation method, based on real-time monitoring systems widely used for earth-rock dams (Zhong et al. 2009, 2011), evaluates compaction quality by monitoring real-time compaction parameters, such as compaction times and running speed, relative to given standards.

However, monitoring technologies fail to establish relationships between the compaction parameters and compactness. Moreover, compaction quality is related to both the compaction parameters and the materials used for constructing dams (Xu and Chang 2013). Many scholars have established compaction quality prediction models on the basis of compaction parameters and material parameters of earth-rock dams in order to control the compaction quality of warehouse surfaces more accurately and comprehensively. Liu et al. (2011) established a multiple regression model to predict compaction quality by using compaction operation parameters and the moisture content of earth-rock aggregate, as obtained from real-time quality monitoring, and analyzed the internal relationship between the compaction parameters and compaction quality (Liu et al. 2014a). In studies on the compactness of other dam types, other compaction models are also presented. Yang and Shi (2010) designed a laboratory experiment to establish a nonlinear regression model between rainfall and the vibrating compacted value (VC value) of roller-compacted cement (RCC) dams. Commuri et al. (2011) constructed a neural network–based intelligent asphalt compaction analyzer to determine the compaction quality of hot asphalt mixes. Lai et al. (2011) evaluated the compaction quality of backfills using stress wave propagation velocities, obtaining a linear relationship between dry densities and stress wave velocities. Liu et al. (2015b), using the climatic conditions of construction and integrating the climatic data and compaction data, established a multiple linear regression model for evaluating the compaction quality of RCC dams.

The aforementioned studies indicate that the compaction parameters, material source parameters, and climatic factors, all affect the compaction quality. However, the attributes involved in the studies are deficient and the precision of the prediction models (e.g., linear regression model, nonlinear regression model, and neural network model) cannot be secured because they each implement empirical
risk minimization (ERM). Recently, some scholars introduced new indices to characterize the compaction quality. Liu et al. (2014b, 2015a) used compaction values (CVs) and unit compaction energy indices to represent the compaction quality; this has been widely used in advanced compaction technology (ACT) in road construction. Tan et al. (2014) used $S_{D}$ (differential of valley values), obtained from the response information of a fiber Bragg grating sensor, to characterize the compaction condition. Kumar et al. (2016) summarized several indices for characterizing dam compactness. Although these indicators increase the real-time accuracy, they are never actually identical to the compactness, and these methods do not comprehensively consider factors that affect the compactness. In addition, during construction, the particle sizes of gravelly soil are widespread, which may produce strong noise in the signals of sensors installed in wheels; therefore, deviations are inevitable when these methods are used in monitoring the compaction.

In order to overcome the deficiencies of the aforesaid investigations, this study designs and implements a multisource heterogeneous data integration subsystem on the basis of a real-time monitoring system to comprehensively consider the attributes that influence the compaction quality. Moreover, in order to improve the accuracy of the compaction forecasting model, a support vector regression (SVR) algorithm and chaos-based firefly algorithm construction model are introduced. Unlike previous models, such as the artificial neural-network model that suffers from local minimum traps and difficulties in determining the hidden layer size and learning rate (Vapnik 1995; Yeh et al. 2011), or other models implementing the ERM principle, SVR, originally introduced by Vapnik (1995), has a global optimum and better prediction accuracy because it uses the structural risk minimization (SRM) principle, which considers both training errors and the capacity of the regression model (Cristianini and Shawe-Taylor 2000; Yeh et al. 2011). Characterized by strong stability and generalization, it can effectively solve nonlinear problems in higher-dimensional spaces and is widely used in different disciplines, such as predicting climatic data under many influencing factors (Kaneda and Mineno 2016), shear strength prediction in reinforced concrete deep beams (Chou et al. 2015), compressive strength prediction of high-performance concrete (Pham et al. 2015), and modeling of temperature-frequency correlations (Hua et al. 2007).

The main problem with SVR is the determination of its hyper-parameters, where selection is based on experiential repeated debugging or filtering by a grid-searching algorithm. Usually, the former is unlikely to obtain the optimal parameter pairs, whereas the latter often consumes too much time and energy. Recently, bionic intelligent algorithms have shown excellent performance in searching for optimal parameters. The firefly algorithm (FA) is a new algorithm in the domain of bionic intelligent algorithms and shows much potential in solving combinatorial optimization problems. The algorithm has been successfully used in solving multireservoir operations in continuous and discrete domains (Garousi-Nejad et al. 2016a), finding the optimal operation of reservoirs for irrigation supply and hydropower production (Garousi-Nejad et al. 2016b), and predicting stock market prices when combined with SVR algorithms (Kazem et al. 2013). However, the original FA uses uniform distributions for population initialization and local exploration; thus, local search ability is limited. Chaos-based methods can successfully replace existing random generators in many applications by increasing the exploration power of the stochastic search process (Fister et al. 2015). The original FA can be enhanced with chaos in two ways. First, the chaotic map can replace some randomly distributed FA parameter in order to improve the performance (Fister et al. 2014; Gandomi et al. 2013). Second, the intrinsic firefly structure can be applied to tune algorithm parameters using a chaotic map (Yang 2015). In this study, the authors propose a chaotic firefly algorithm (CFA) to optimize the SVR hyperparameters, and use the optimized SVR to reconstruct the compaction quality prediction model. The results show that our model performs better than previous models. The authors also compare the proposed chaos-based FA to several different bionic intelligent algorithms to demonstrate its practicability and effectiveness.

This paper is organized as follows: section “Methodology” provides a brief introduction to SVR and FA, and the method of SVR optimization by chaotic FA is proposed. Section “Case Study on Compaction Quality of Earth-Rock Dam” consists of four parts. The first part briefly introduces the real-time construction-quality monitoring system and the dam construction personal digital assistant (PDA) collection system, as well as data structure descriptions of these two systems. The second part focuses on the design and implementation of a data integration framework according to data characteristics. The third part constructs the compaction quality prediction model using the proposed method. This optimization strategy is compared to the classical bionic intelligent algorithm of particle swarm optimization (PSO) and the new bionic intelligent algorithm of bacterial foraging (BF) to prove its practicability and effectiveness. The authors also demonstrate that this prediction model has significantly improved practicability, effectiveness, and precision compared with the current commonly used models. In the fourth part, a whole storehouse of compaction quality is evaluated by this model. The conclusions of this study and future research are discussed in the final section.

Methodology

Introduction to SVR and FA Algorithms

Support vector regression is the use of a support vector machine (SVM) in the domain of functions (Vapnik 1999), which is mainly used in approximating functions. Assuming that the fit line of SVR is

$$y = [\omega]^{T}\phi([x]) + b \quad (1)$$

where $\phi([x])$ = nonlinear mapping of data in higher-dimensional feature space, and $[\omega]$ and $b$ are functionally related (in this paper, $[x]$ refers to the parameter vector that affects the compactness, and $y$ refers to the compactness); $[\omega]$ and $b$ can be determined on the basis of the consistency principle of SRM. To avoid overfitting, a slack variable is introduced. Therefore, the SVR fitting problem can be demonstrated as follows:

$$\text{minimize} \quad \frac{1}{2}[\omega]^{T}[\omega] + C \sum_{i=1}^{n}(\xi_{i} + \xi_{i}^{*}) \quad (2)$$

subjected to

$$\begin{cases} y_{i} - \omega^{T}[x]_{i} - b \leq \varepsilon + \xi_{i} \\ \omega^{T}[x]_{i} + b - y_{i} \leq \varepsilon + \xi_{i}^{*} \\ \xi_{i}, \xi_{i}^{*} \geq 0, \quad i = 1, 2, \ldots, n \end{cases}$$

where $C \geq 0 = \text{penalty factor}; \xi_{i} \geq 0 = \text{slack quantity above the upper bound}; \xi_{i}^{*} \geq 0 = \text{slack quantity below the lower bound};$ and $n = \text{number of samples}. \text{From Eq. (2), it is easy to conclude that the specific form of SVR can be determined only if the proper penalty parameter C, insensitive loss coefficient} \varepsilon,$ and kernel function are selected; these are crucial parameters influencing the
performance of SVR. In actual operation, the selection of parameters is based on experiential repeated debugging or grid search algorithms.

The FA was proposed by Yang (2010) and realizes the optimization of target functions through simulating the attraction and motion processes of firefly clusters. Fireflies with brighter fluorescence have stronger appeal; eventually, many fireflies gather around the fireflies with the brightest fluorescent elements. The position upgrading of the ith firefly attracted by and moving toward the jth firefly can be determined by Eq. (3)

$$p_i = p_i + \beta \times (p_j - p_i) + \alpha \times (\text{rand} - 0.5)$$  

where $p_i$ and $p_j$ are positions of the ith and jth fireflies; $\alpha$ is a step factor; and rand is random factor evenly distributed in [0, 1]. This disturbance term is added to avoid situations in which the algorithm prematurely falls into local optimization.

The concrete operational procedure of the algorithm runs as follows:

1. Randomly distribute the firefly group in the solution space.
2. Calculate the fluorescent brightness on the basis of the positions of the fireflies. By comparing relevant fluorescent brightness values $I$, fireflies with higher fluorescent brightness attract those with lower fluorescent brightness to move toward them using Eq. (3).
3. Repeat Step 2 until the maximum iterations are reached.
4. The solution to the target function is attained by the position of the brightest firefly.

**Realization and Application of the SVR with CFA**

To draw a conclusion, the complexity and generalization of the SVR algorithm. The concrete operational procedure of this algorithm runs as follows:

1. Logistic mapping is used to generate initial positions of fireflies.
   - Map the search interval into the range of [0, 1]
     $$x_n^{(l)} = \frac{x_n^{(l)} - x_{\text{min}}^{(l)}}{x_{\text{max}}^{(l)} - x_{\text{min}}^{(l)}}$$  

where $n = $ serial number of the nth firefly; $x_{\text{max}}^{(l)} = $ maximum value of the ith dimension; and $x_{\text{min}}^{(l)} = $ minimum value of the ith dimension.
   - Adopt logistic mapping to generate the next iteration
     $$x_{n+1}^{(l)} = \mu x_n^{(l)}(1 - x_n^{(l)})$$  

   - Map the chaotic sequence into the search space
     $$x_{n+1}^{(l)} = x_{\text{min}}^{(l)} + (x_{n+1}^{(l)} - x_{\text{min}}^{(l)})$$  

2. Calculate the brightness of each firefly. In this study, the number related to the mean square error (MSE) is adopted as the fitness function. The formulation of MSE is as follows:

$$\text{MSE} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y_i - f_i)^2}$$

where $f_i = $ actual value and $y_i = $ predicted value.

3. Update the position of each firefly.
   - Use logistic mapping to generate a chaotic sequence.
     $$x_n^{(l)} = \mu x_n^{(l)}(1 - x_n^{(l)}) \quad n = \{2, 3, \ldots, K\}$$

where $K = $ length of chaotic sequence and $c_1^{(l)} = $ evenly distributed random number between (0, 1).
   - Use a chaotic component, $\lambda(c_K^{(l)} - 0.5)$ instead of random component $\alpha \times (\text{rand} - 0.5)$, where $\lambda$ is the contraction factor and expressed as follows:

$$\lambda = \frac{\text{MI} - l + 1}{\text{MI}}$$

### Table 1. Test Results of Benchmark Functions

<table>
<thead>
<tr>
<th>Number</th>
<th>Exact value</th>
<th>Search value</th>
<th>Fitness</th>
<th>Convergence times</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>(0.0)</td>
<td>$(-2.60 \times 10^{-6}, -1.22 \times 10^{-5})$</td>
<td>$0$</td>
<td>$3.542 \times 10^{-05}$</td>
</tr>
<tr>
<td>$f_2$</td>
<td>(0, -1)</td>
<td>$(1.13 \times 10^{-5}, -0.9999997)$</td>
<td>$0$</td>
<td>$3.000000002$</td>
</tr>
<tr>
<td>$f_3$</td>
<td>(1.1)</td>
<td>$(0.9999958, 1.000016)$</td>
<td>0</td>
<td>$1.817 \times 10^{-09}$</td>
</tr>
<tr>
<td>$f_4$</td>
<td>$(512, 404.2319)$</td>
<td>$(512, 404.2318045125)$</td>
<td>$-959.6407$</td>
<td>$-959.6406627$</td>
</tr>
<tr>
<td>$f_5$</td>
<td>$(0.125313)$</td>
<td>$(-2.1893, 0.28499)$</td>
<td>$0.292579$</td>
<td>$0.293940426$</td>
</tr>
<tr>
<td>$f_6$</td>
<td>$(1.111)$</td>
<td>$(1.00110021004)$</td>
<td>0</td>
<td>$4.012433 \times 10^{-05}$</td>
</tr>
<tr>
<td>$f_7$</td>
<td>$(0.00)$</td>
<td>$(2.5 \times 10^{-5}, 2.5 \times 10^{-5}, -4 \times 10^{-5})$</td>
<td>0</td>
<td>$5.767 \times 10^{-07}$</td>
</tr>
</tbody>
</table>

Note: The specific expressions from $f_1$ to $f_7$ are given in the Appendix.

where $M_I = \text{maximum number of iterations of the algorithm}$ and $t = \text{current iteration number of the algorithm}$.

4. Repeat Step 2 until the maximum iterations are reached. Upon terminating the algorithm, the position of the brightest firefly is used as the input parameters for the SVR algorithm.

To prove the practicability and effectiveness of the proposed search strategy, several benchmark functions are adopted to test its performance.

According to Table 1, the CFA algorithm proposed in this research is practical and effective. Therefore, the algorithm can be used to optimize the hyperparameters of SVR. The central idea of the SVR with CFA runs as follows: define the parameter sets ($\gamma$, $\varepsilon$, $C$) as the search space of the fireflies and allow the fireflies to move toward the optimal value by the mutual degree of attraction, thereby determining the optimal positions of fireflies and setting these as the input parameters for the SVR algorithm. The object-oriented language C# is adopted to realize this algorithm. A pseudocode of the algorithm is illustrated in Algorithm 1.

Algorithm 1. Pseudocode of Support Vector Regression with Chaos-Based Firefly Algorithm

**Input:** Population of fireflies $X = \{X_1, X_2, \ldots, X_n\}$, fitness function $f(X_i)$.

**Output:** The position of the brightest firefly.

1. Logistic mapping is used to generate initial positions of fireflies:
   $$X^{(0)} = \{X_1^{(0)}, X_2^{(0)}, \ldots, X_n^{(0)}\}.$$
2. for $k = 1$: Max Iteration Number
3. for $i = 1$: Pop_Size
4. for $j = 1$: Pop_Size
5. if $FA_i \cdot \text{Intensity} < FA_j \cdot \text{Intensity} \&\& ||FA_i \cdot \text{Position} - FA_j \cdot \text{Position}|| < \text{Radius of view}$ then
6. Chaotic movement of the ith firefly using Eq. (13).
7. end if
8. end for
9. $SVR_{Modelk} = SVR \cdot \text{Train} (FA_i \cdot \text{Position}, \text{Training Data})$
10. $FA_i \cdot \text{Intensity} = SVR_{Modelk}, \text{Accuracy} = SVR \cdot \text{Predict} (SVR_{Modelk}, \text{Test Data})$
11. end for
12. Rank fireflies and find the best one
13. end for

In Algorithm 1, $\text{Max Iteration Number}$ represents the number of maximum iterations and $\text{Pop Size}$ represents the size of the firefly group, $SVR_{Train} (FA_i \cdot \text{Position}, \text{Training Data})$ means that the positions of the fireflies are used as the input parameters of SVR and the training data is used for the training of the model, $SVR_{Predict} (SVR_{Modeli}, \text{Test Data})$ means that the training data is used to test the model. Smaller values of $\text{MSE}$ returned from the model correspond to brighter fireflies, and $\text{Radius of view}$ indicates the radius of the visual field of the fireflies.

In this study, the proposed algorithm is used for compaction quality evaluation, which is convenient for real-time guidance. The main steps are as follows (Fig. 1): First, a multisource heterogeneous data integration subsystem is designed to integrate compaction monitoring data, material source statistical data, and detected data of test pits from the historical database, as detailed in the next section. Then, SVR with CFA is used to establish the compaction quality prediction model. Finally, the model is used to evaluate the compaction quality of a storehouse surface.

**Case Study on Compaction Quality of Earth-Rock Dam**

In this section, the authors test the proposed algorithm using the monitoring data of the Changhe dam, a large-scale high earth-rock dam in the southwest of China. Two different systems were used to capture the heterogeneous data, and a data integration subsystem for three different types of data was designed and implemented. Then the proposed algorithm was tested on the integrated data. Some optimization strategies are compared with the CFA, and the compaction quality prediction model reconstructed by SVR with CFA is also compared with current models based on ERM.

**Data Collection**

**Real-Time Monitoring System of Construction Quality and Data Description**

The real-time monitoring system of construction quality (RMS) uses technologies of global positioning systems (GPS), computer networks, and automatic control to realize the real-time collection of compaction data. The structure of the system is shown in Fig. 2.

Data collected by the mobile station mainly included compaction monitoring data, material source statistical data, and travel speeds. This system can generate the original images of

---

**Fig. 1. (Color) Application of SVR with CFA algorithm in compaction quality evaluation**
the rolling thickness and rolling passes, which are later uploaded and stored in the file transfer protocol (FTP) server.

Personal Digital Assistant Collection System for Dam Construction Data and Data Description
Information of the construction site, including the rolling machine information, material source information, and spot test information, were collected by a PDA collection system (PCS). These data could be sent back to the server through the PDA, or uploaded to the server via the web client entry, as shown in Fig. 3.

Material source and spot tests information mainly included the location information (dam coordinates) of samples, design information of material sources, particle gradation and content information of various materials, and quality inspection information. These data were stored in the web database for subsequent quality inspections.

Data Integration
As noted previously, the dynamic position information, speed, and vibrational states of the rolling machine collected by the RMS were stored in the compaction database. The FTP server mainly stored the original altitude image of altitude data stored by the gray value, which is generated by RMS. The material source data and compaction quality data were input by PDA or web client entry and stored in the web database. Because these three different types of data have variant sources without directly related key links available, an integration framework for the three heterogeneous data types was designed to integrate these three types of heterogeneous data.

There are three sections in the process of data integration: coordinate conversion of compaction data, filtering of compaction data and graphic data on the basis of position information, and the generation of compaction parameters and integrating these with material source data and quality data. In order to realize the framework of data integration, the object-oriented language C# was used to establish a subsystem of data integration. The flowchart of the algorithm is shown in Fig. 4.

This study used the quality inspection data of the core wall area from August 2, 2013, to September 10, 2015, as the integration
object. In order to integrate the quality inspection data with compaction data, each quality inspection data set processed 70,000–80,000 compaction data points for a storehouse surface. After removing all compaction data that could not be integrated because of power failures, network outages, or signal loss, 620 data points were integrated. These consisted of 27 attribute values, including position information (horizontal coordinate, vertical coordinate, and altitude), material source and quality inspection data (design parameters, maximum particle size, content of various), and compaction parameter data (rolling thickness and rolling passes), as shown in Table 2.

### Establishment of Compaction Quality Prediction Model Based on SVR with CFA

The integrated data of compaction quality contains a large number of attributes which may be either useful or redundant. Redundant attributes may affect the efficiency of the algorithm and may lead to a deviation of results. Therefore, it is necessary to carry out the effective data reduction. In this study, compactness of the whole material was regarded as a criterion for the compaction quality. Whereas the relationship between the source material parameters and compaction parameters was needed, other compaction quality parameters (i.e., the whole material dry density, the fine material dry density, the fine material compaction, the maximum dry density of the whole material, and the maximum dry density of the fine material) were ruled out. This study used the data mining platform weka and adopted Boosting Tree and Genetic Search algorithms to find out the important attribute. The authors searched the integrated data from the core wall area to find the important attributes in the following manner. First, open the weka platform, select Applications Explorer to open weka explorer, and select Open file to load the data. Then select the Select attributes function, choose the algorithm in the Search method, and click on Start button to start the operation. The core idea of Boosting Tree is to create many trees, such as binary trees, to split the training data and approximate to a given vector function. The importance of each attribute is measured by their contribution to the precision of the training data. The genetic search algorithm uses particular evaluation methods to choose the best attribute population or the best combination of attribute populations from those representing a subset of attributes. Finally, nine attributes, namely maximum particle size, P5 content, moisture content, organic content, plastic index as well as rolling velocity, rolling thickness, high vibration passes, and static rolling passes were chosen to evaluate the compactness of blended material.

Based on the algorithm flowcharts in Algorithm 1, the authors randomly distributed the 620 data points which were integrated, partly used for training and partly for testing. To elevate the generalization of the SVR algorithm and to eliminate overfitting, the distribution of the data set is vitally important apart from controlling the input parameters. Obviously, the authors did not simply use...
the data set to train the model to make the model a better fit for the training data set. Poor performance for the test data indicated that overfitting was happening. Inversely, more data sets should be taken for testing to improve the generality of the model. Therefore, before training, the aforementioned integrated data were randomly divided into two sets: 50% used for training and the other for testing. Based on the research introduced by Üstün et al. (2005), the parameters of SVR algorithm were set to $C = 10^8$, $\varepsilon = 0.01$, and $\gamma = 0.1$. To ensure that all parameters were of the same order of magnitude, $C$ was confined within the range of $[1,200]$, $\varepsilon$ multiplied by 1,000 to provide locations in the range of $[0,200]$, and $\gamma$ multiplied by 100 to enlarge it to $[1,200]$, so that the FA could search in a three-dimensional space similar to a cube. The size of the firefly group was set to 50. To avoid fireflies blindly relying upon the optimal current individual to lengthen the convergence time, the radius of the visual field of the fireflies was set to $1/4$ of the longest distance of the search space. Generally speaking, the maximum degree of attraction and maximum light intensity coefficient were set to 1. And the fitness function was set as follows:

$$
\text{Fitness} = e^{R_1 - R_2 \cdot \text{MSE}}
$$

where $R_1$ and $R_2$ are adjustable parameters; and MSE = mean square error of the model. In this study, $R_1$ was set as 10, and $R_2$ was set as 5.

At the beginning of the algorithm, the fireflies, whose brightness values are related to the MSE of the test set, were distributed in the search space using logistic mapping. Smaller MSE values correspond to brighter fireflies. Subsequently, with iterations of the algorithm, the fireflies move toward the optimal position. When the algorithm reaches the maximum number of iterations, the algorithm terminates and the position of the optimal firefly is selected as the input parameter of the SVR algorithm. Fig. 5 shows the initial positions of the fireflies and Fig. 6 displays the movement of fireflies.

To prove the practicability and effectiveness of the algorithm proposed in this research, the algorithm is compared to a basic FA, a classic bionic intelligent algorithm of PSO, and a new bionic intelligent algorithm of BF. Table 3 shows the conclusions of the different algorithms. Fig. 7 is the convergence situations of these algorithms.

From Table 3, it is easy to conclude that the CFA is operable and efficient. Compared with the other three algorithms, the proposed algorithm has a higher convergence rate, higher accuracy, and higher robustness. It should be noted that the optimized results only apply to the integrated data set in this study; for other data set, the applicable parameters can be found according to the algorithm flow.

In order to prove the precision of the compaction quality prediction model constructed by the proposed algorithm in this research, the predicted outcomes of this algorithm are compared with those of the multiple linear regression and neural network model, as shown in Fig. 8. The linear correlation coefficient $R$ [Eq. (12)], the mean absolute error MAE [Eq. (13)], the relative absolute error RAE [Eq. (14)], the mean square error MSE [Eq. (8)], the maximum error MAX, and the minimum error MIN are adopted as accuracy criteria.

<table>
<thead>
<tr>
<th>Table 3. Results of Various Optimization Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
</tr>
<tr>
<td>$C$</td>
</tr>
<tr>
<td>$\gamma$</td>
</tr>
<tr>
<td>$\varepsilon$</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>Convergence times</td>
</tr>
</tbody>
</table>

Note: The data in the table is the average result of each algorithm independently running 20 times.
\[ R = \frac{N \sum \text{AMC} \cdot \text{AMC} - (\sum \text{AMC})^2}{\sqrt{N(\sum \text{AMC}^2) - (\sum \text{AMC})^2}} \]

In Eqs. (12)–(14), \( \text{AMC} \) is the predicted value; \( \text{AMC} \) is the actual measurement of the compactness; \( \overline{\text{AMC}} \) represents the mean of the actual value; and \( N \) = number of data.

It can be seen from Fig. 8 and Table 4 that the proposed algorithm is practicable, as well as more effective and accurate compared with the multiple linear regression model and neural network model. Therefore, this model can be used in the evaluation of compaction quality.

**Evaluation of Work Unit Compaction Quality**

The monitoring data of a work unit from the Changhe dam is used for the evaluation of compaction quality. As discussed previously, the compaction data was collected once per second to record the position information, vibration state information, and velocity information. Based on the PDA real-time input system, the material source data of the storehouse surface and the quality inspection data were collected and stored in the web database. To begin, the newly established integrated data subsystem was used to generate the compaction parameters of every point (every pixel point of the surface in screen coordinates). Then the surface was partitioned with a 0.3 × 0.3-m lattice and the average value of the data in each lattice unit was used to represent the compaction parameters (average velocity,
Table 4. Error Statistic of Different Models

<table>
<thead>
<tr>
<th>Error measures</th>
<th>Multiple linear regression model</th>
<th>Neural network model</th>
<th>SVR with CFA model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.908396</td>
<td>0.935089</td>
<td>0.960455</td>
</tr>
<tr>
<td>MAE</td>
<td>0.307064725</td>
<td>0.301873786</td>
<td>0.202913637</td>
</tr>
<tr>
<td>RAE</td>
<td>0.357421381</td>
<td>0.351379162</td>
<td>0.236190179</td>
</tr>
<tr>
<td>MSE</td>
<td>0.423149</td>
<td>0.41563913</td>
<td>0.258316313</td>
</tr>
<tr>
<td>MAX</td>
<td>2.632</td>
<td>2.215</td>
<td>0.92748</td>
</tr>
<tr>
<td>MIN</td>
<td>0</td>
<td>0.001</td>
<td>0.000799192</td>
</tr>
</tbody>
</table>

Note: The multiple linear regression (MLR) model is $AMC = 0.0064 \times MPS + 0.0636 \times PSC + 0.0438 \times AMMC - 9.8125 \times OC + 0.0625 \times PI - 0.3768 \times AV - 2.1576 \times Thic + 0.0557 \times HVC + 0.2561 \times NVC + 95.6512$.

The quality evaluation of widely used earth-rock dams is of vital significance. As a result, the intelligent integration of compaction data and material source data, as well as the establishment of a quality prediction model using these data, is among the most significant and difficult points of current research. This research designs and realizes a heterogeneous data integration framework for three different sources of data. On this basis, a high-precision compaction quality prediction model based on SVR with CFA is proposed. The achievements of this research are as follows:

1. To establish a high-precision prediction model, data integration is an essential link. The different systems used in earth-rock dam quality monitoring create different data forms. As a result, the integration of heterogeneous data is different in actual operation. By using object-oriented technology, this study integrates three different types of data from different sources and of variant forms, which provides a data source for the establishment of a high-precision prediction model.

2. This study proposes SVR with CFA to create the prediction model. That is to say, the CFA is used to optimize the input parameters of the SVR algorithm, which is used to establish the compaction quality prediction model. Compared with the multiple linear regression model and neural network frameworks that are currently widely used, this model shows significant improvements in prediction precision, which corroborates the feasibility of the algorithm.

3. When used in predicting the quality of storehouse surface compaction, the proposed algorithm reveals the surface compaction quality more comprehensively than traditional tests of test pits and graphical reports of compaction parameters can show.

This study represents a preliminary exploration of compaction quality evaluations that combine bionic intelligent algorithms and data mining algorithms, in the hope that construction workers can take control of compaction quality outcomes more comprehensively and accurately. If this algorithm is embedded in real-time monitoring systems for compaction, its real-time guiding advantages will be apparent.

### Conclusion and Prospect

The quality evaluation of widely used earth-rock dams is of vital significance. As a result, the intelligent integration of compaction data and material source data, as well as the establishment of a quality prediction model using these data, is among the most significant and difficult points of current research. This research...
5. Schaffer function \( N \) 4
\[
f_5 = 0.5 + \cos^2(\sin(|x_1^2 - x_2^2|)) - 0.5 \left(1 + 0.001(x_1^2 + x_2^2)^2 \right)\cos(x_1^2 - x_2^2),
\]
x \in [-100, 100], \ i = 1, 2

6. Rosenbrock function
\[
f_6 = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]
\]

7. Rastrigin function
\[
f_7 = A n + \sum_{i=1}^{n} [x_i^2 - A \cos(2\pi x_i)],
\]
where \( A = 10 \), \( x_i \in [-5.12, 5.12] \)

Acknowledgments

The authors wish to thank the Digital Dam project department of the CHB hydroelectric station for providing the data necessary to complete this study. This research was funded by the Innovative Research Groups of the National Natural Science Foundation of China (51321065), the National Basic Research Program of China (973 Program) (2013CB035904), and the National Basic Research Program of China (973 Program) (2013CB035906).

References


Weka [Computer software]. Univ. of Waikato, Hamilton, New Zealand.