Advances in Engineering Software 65 (2013) 182–190

Contents lists available at SciVerse ScienceDirect

Advances in Engineering Software

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

Adaptive system for dam behavior modeling based on linear regression and genetic algorithms

B. Stojanovic [a,*], M. Milivojevic [b], M. Ivanovic [a], N. Milivojevic [c], D. Divac [c]

[a] Faculty of Science, University of Kragujevac, Radoja Domanovica 12, 34000 Kragujevac, Serbia
[b] Technical and Business College, Trg Sv. Save 34, 31000 Uzice, Serbia
[c] "Jaroslav Cerni" Institute for the Development of Water Resources, 11000 Belgrade, Serbia

ARTICLE INFO

Article history:
Received 25 February 2013
Received in revised form 15 May 2013
Accepted 26 June 2013
Available online 20 July 2013

Keywords:
Dam
Modeling
Linear regression
Genetic algorithms
Adaptive system
Monitoring

ABSTRACT

Most of the existing methods for dam behavior modeling require a persistent set of input parameters. In real-world applications, failures of the measuring equipment can lead to a situation in which a selected model becomes unusable because of the volatility of the independent variables set. This paper presents an adaptive system for dam behavior modeling that is based on a multiple linear regression (MLR) model and is optimized for given conditions using genetic algorithms (GA). Throughout an evolutionary process, the system performs real-time adjustment of regressors in the MLR model according to currently active sensors. The performance of the proposed system has been evaluated in a case study of modeling the Bocac dam (at the Vrbas River located in the Republic of Srpska), whereby an MLR model of the dam displacements has been optimized for periods when the sensors were malfunctioning. Results of the analysis have shown that, under real-world circumstances, the proposed methodology outperforms traditional regression approaches.

1. Introduction

Dams have strong interactions with environmental, hydraulic and geomechanical factors, such as air and water temperature, water level, pore pressure, and rock deformability, each of which influences the structural behavior of the dam [1]. To describe and predict the structural behavior of dams, over the past decades, a number of deterministic, statistical and hybrid mathematical models have been developed.

For a long time, statistical models have been applied to dam safety monitoring to find out the contribution of external loads to dam displacements [2]. A number of statistical models based on multiple linear regression (MLR) and their advanced forms such as hierarchical regression, stepwise multiple regression, robust regression, ridge regression, and partial least squares regression have been shown to be more or less successful in dam modeling [3]. The advantages of the statistical models are the simplicity of formulation, the speed of execution and the availability of any type of correlation between independent and responses variables. In contrast, deterministic models require solving differential equations, for which closed form solutions could be difficult or impossible to obtain [4]. Therefore, many models that are based on numerical methods, such as the finite element method, have been developed as well [5]. Recently, numerical and statistical methods have been enriched with various heuristics from the artificial intelligence (AI) domain, creating hybrid models that combine their advantages. Some of these artificial intelligence techniques and heuristic algorithms are artificial neural networks (ANN) [6], genetic algorithms (GA) [7], support vector machines (SVM) [8], adaptive neuro-fuzzy inference systems (ANFIS) [9,10], Monte Carlo simulations [11], the modified complex method [12] and the artificial immune algorithm (AIA) [13].

Recently, Rankovic et al. [9] presented a study in which the objective was to develop an adaptive neuro-fuzzy inference system (ANFIS) to predict the radial displacements of Bocac arch dam. ANFIS models have been proposed as an alternative approach for an evaporation estimation of the Yuvacik Dam [10]. In his paper, Mata [6] presented a comparison between MLR and ANN models for the characterization of dam behavior under environmental loads for the Alto Rabagao arch dam. In their study, Wang et al. [14] investigated several Artificial Intelligence techniques for modeling monthly river flow discharge time series, which included an ANN approach, an ANFIS technique, GP models and support vector machines (SVM), and compared their performance with traditional time series modeling techniques, such as autoregressive moving-average (ARMA) models. To improve prediction, support vector regression (SVR) upgraded with GAs is often combined with existing ANN and ARMA techniques [8]. Hybrid algorithms and their

* Corresponding author. Tel.: +381 69 11 543 75.
E-mail addresses: bobi@kg.ac.rs (B. Stojanovic), milovan.milivojevic@vpts.edu.rs (M. Milivojevic), mivanovic@kg.ac.rs (M. Ivanovic), nikola.milivojevic@gmail.com (N. Milivojevic), ddivac@eunet.rs (D. Divac).

0965-9978/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.advengsoft.2013.06.019

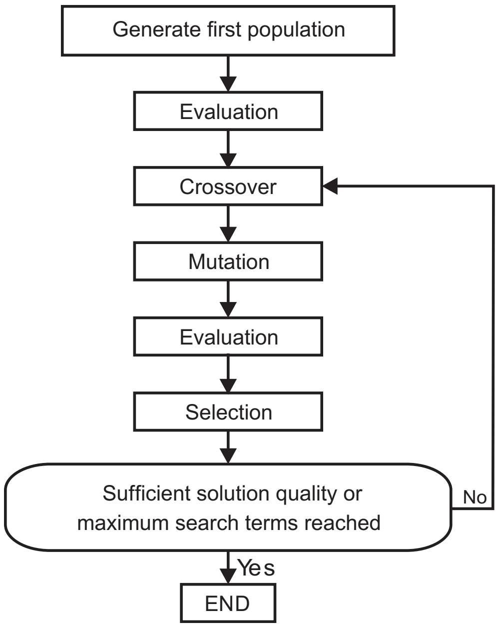

Fig. 1. Schematic view of a genetic algorithm.

applications in prediction and optimization are also presented in a number of papers. Gholizadeh and Seyedpoor [15] used a hybrid methodology with a combination of metaheuristics (GA and Particle Swarm Optimization-PSO) and neural networks to propose an efficient soft computing approach to achieve optimal shape design of arch dams that were subjected to natural frequency constraints. There are several recent studies that describe applications of AIA techniques, which imitate the function of the natural immune system. For example, Xi et al. [16] proposed an immune statistical model to resolve the data analysis problems of dam horizontal crest upstream–downstream displacements.

The above-mentioned methods play a crucial role in modeling structural dam behavior. However, all of the methods require a persistent set of input parameters, i.e., all of the measurements must be available during the entire examined period. In real-world applications, failures of the measuring equipment can lead to a situation in which the selected model becomes unusable because of volatility of the independent variables. In this paper, the concept of an adaptive system for dam behavior modeling (ASDBM) that is resistant to the variations in the set of measured variables is presented. This system is based on a multiple linear regression model of a dam, the form of which is optimized for the given conditions using genetic algorithms.

2. Theoretical background

The adaptive system for the generation of optimized models of dam behavior is a hybrid solution that is based on linear regression models and genetic algorithms. The theoretical background of the applied methods is given in the following sections.

2.1. Multiple linear regression

An MLR is a method that is used to model a linear relationship between a dependent variable (a predictant) and one or more independent variables (predictors). The general form of an MLR can be written as follows:

$$y_i = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \cdots + \beta_j \cdot x_{ji} + \cdots + \varepsilon_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where y_i is a response variable, x_{ji} the predictors ($j = 1, 2, \ldots, k$), k the number of significant predictors, the index i shows the sample number, and ε_i is the independent and normally distributed random variables that have a mean of zero and a variance σ^2. In the majority of the applications of linear regression models, the functional forms of the predictors (basis functions) are not clear in advance and are dependent on the nature of the modeled phenomenon [17]. Coefficients $\beta_0, \ldots, \beta_k$ are unknown parameters of the model, which are estimated for a given set of data using the least squares method, which minimizes the sum of the squared error (SSE):

$$SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - b_0 - \sum_{j=1}^{k} b_j \cdot x_{ji}\right)^2 \tag{2}$$

by taking the derivatives of SSE with respect to the b's and setting them equal to zero. In Eq. (2), the b's are the estimated values of the model parameters.

2.2. Genetic algorithms

Genetic algorithms are search techniques that are inspired by the theory of natural selection, in which strong species have a greater opportunity to survive and pass their genes onto future generations via reproduction [18]. GAs are probabilistic algorithms that maintain a population of individuals, $P(t) = x_1(t), \ldots, x_n(t)$, for the iteration (generation) t, where each individual represents a potential solution to the problem. Each solution $x_i(t)$ is evaluated to quantify its fitness. Subsequently, a new population (iteration $t + 1$) is formed by selecting better individuals from those of generation t. In GA terminology, a solution vector $\mathbf{x} \in \mathbf{X}$ is called an individual or chromosome and corresponds to a unique solution $\mathbf{x}$ in the solution space. The chromosomes are made of discrete units called genes. Each gene controls one or more features of the chromosome. In this paper, genes are assumed to be binary digits, according to the original implementation of GAs by Holland.

GA uses two operators to generate new solutions from the existing solutions: crossover and mutation. In crossover, two chromosomes, called parents, are combined together to form new chromosomes, called offspring. By iteratively applying the crossover operator, genes of good chromosomes are expected to appear more frequently in the population, which eventually leads to convergence to an overall optimal solution. The mutation operator introduces random changes into the characteristics of chromosomes, for the purpose of reintroducing genetic diversity back into the population and assisting the search to escape from local optima.

Reproduction involves selecting a set of chromosomes that will survive into the next generation. There are different selection procedures in GAs depending on how the individual fitnesses are used. Proportional selection, ranking, and tournament selection are the most popular selection procedures. The procedure of a generic GA [19] is given in Fig. 1.

After the random generation and evaluation of initial solutions, the population is subjected to the iterative process of mating (crossover), mutation, evaluation and selection for the next iteration (generation). The iterative process is terminated when there is a satisfactory quality of the solutions or when a maximum number of iterations is reached.

3. Adaptive system for dam behavior modeling

Among the main problems in real-time applications of predictive models are irregular operations of the measuring equipment, which lead to an impermanent set of available measured variables. The goal of the developed ASDBM system is to achieve an automatic real-time generation of the optimized predictive regression model that is robust to possible changes in the set of input parameters.