Structural Safety 48 (2014) 33–39

Contents lists available at ScienceDirect

Structural Safety

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

Development of support vector regression identification model for prediction of dam structural behaviour

Vesna Ranković [a,*], Nenad Grujović [a], Dejan Divac [b], Nikola Milivojević [b]

[a] Department for Applied Mechanics and Automatic Control, Faculty of Engineering, University of Kragujevac, Sestre Janjić 6, 34000 Kragujevac, Serbia
[b] Institute for Development of Water Resources "Jaroslav Černi", 80 Jaroslava Černog St., 11226 Beli Potok, Belgrade, Serbia

ARTICLE INFO

Article history:
Received 30 September 2013
Received in revised form 14 December 2013
Accepted 11 February 2014
Available online 18 March 2014

Keywords:
Dam
Nonlinear behaviour
Identification
NARX
Support vector regression
Tangential displacement

ABSTRACT

The paper presents the application of support vector regression (SVR) to accurate forecasting of the tangential displacement of a concrete dam. The SVR nonlinear autoregressive model with exogenous inputs (NARX) was developed and tested using experimental data collected during fourteen years. A total of 573 data were used for training of the SVR model whereas the remaining 156 data were used to test the created model. Performance of a SVR model depends on a proper setting of parameters. The SVR parameters, the kernel function, the regularization parameter and the tube size of ε-insensitive loss function are specified carefully by the trail-and-error method. Efficiency of the SVR model is measured using the Pearson correlation coefficient (r), the mean absolute error (MAE) and the mean square error (MSE). Comparison of the values predicted by the SVR-based NARX model with the experimental data indicates that SVR identification model provides accurate results.

1. Introduction

Dam parameters monitoring through installed instrumentation is the most important part of a dam safety program [1]. These parameters include seepage flows, seepage water clarity, pore pressure, deformations or movements, water levels, pressures, loading conditions, temperature variations, etc. Physical interpretation of significant indicators of the structural behaviour is the key factor to control and management of the dam system.

Structural health monitoring of dams is based on acquisition of displacement measurements [2]. Deformation monitoring can reflect the structural behaviour of the dam [3]. Timely and accurate analysis and prediction of the dam displacement is an essential part of the dam safety control.

The structural response of the dam is affected by many factors including reversible (hydrostatic pressure and temperature) and irreversible factors (due to residual deformations associated with creep, alkali-aggregate reaction and other nonlinear effects that may jeopardize the structural integrity [4]). There are different approaches to developing models for prediction of the nonlinear structural behaviour of the dam, and they include deterministic, statistical and hybrid models, which combine the first two. Deterministic modelling requires solving nonlinear partial differential equations for which closed form solutions may be difficult or impossible to obtain [5]. As a result, numerical methods, such as the finite element method, the finite-difference method and the finite volume method are employed. Advantages of the statistical methods include simplicity of formulation, speed of execution, availability of any kind of correlation between independent and dependent variables [6,7]. Performance of the existing statistical regression models is not satisfactory when multicollinearity and influential outliers exist between the variables [8].

Artificial intelligence techniques such as artificial neural networks, fuzzy logic systems, neuro-fuzzy systems and genetic algorithms have been used as effective alternative tools for modelling of complex civil engineering systems and they have been widely used for prediction and forecasting. Furthermore, in dam engineering, these techniques have been successfully used to obtain the optimal shape of dams [9,10] as well as for modelling of the dam behaviour [11–14].

System identification of dams is a significant field of structural engineering [15]. Displacement of the dam is a nonlinear time-varying function of hydrostatic pressure, temperature and other unexpected unknown causes. Nonlinear black-box system identification can be applied to develop complex nonlinear models. NARX input–output model can be used to describe nonlinear structural behaviour of the dam [14]. The output of the NARX model depends on the previous values of itself and inputs. Determination of the model order and the model structure of a general NARX model is

* Corresponding author. Tel.: +381 34 335 990; fax: +381 34 333 192.
E-mail address: vesnar@kg.ac.rs (V. Ranković).

http://dx.doi.org/10.1016/j.strusafe.2014.02.004
0167-4730/

a difficult task even for a single-input and single-output system [16]. Selection of a near-optimal set of time lags is an important and difficult computational task [17]. The main problem in model identification is to approximate the unknown function within a given accuracy from some sampled data sequences. This function is approximated by some general function approximators such as neural networks [18] or neuro-fuzzy systems [19,20].

SVR has recently been used in the framework of the nonlinear black-box system identification [21–24]. In the present work, SVR is used for structural identification of the dam. The support vector machine (SVM) is a new technique that has been intensively used to solve pattern classification and function–approximation problems in many areas. Hammer and Gersmann [25] showed the universal approximation capability of SVMs with various kernels, including Gaussian, several dot products, or polynomial kernels. The SVM implements the structural risk minimization principle, which has been shown to be superior to the traditional empirical risk minimization principle implemented by most of the conventional neural networks. Training of the SVM is equivalent to solving a linearly constrained quadratic programming problem so that the solution of the SVM is globally optimal and unique [26]. The support vector machine is based on the statistical learning theory [27]. The SVM was originally developed for binary classification problems. Support vector regression (SVR) presents an application of the SVM for function estimation [28,29].

The procedure based on neuro-fuzzy modelling was presented and discussed for the radial displacement of an arch dam by Rankovic et al. [30].

The objective of this study is to develop a support vector regression-based NARX model for the dam tangential displacement prediction and to demonstrate how it is applied to identify complex nonlinear relationships between the input and output variables.

2. Case study

Construction of the hydropower plant Djerdap 2 officially began in 1977, and the first units were put into operation in 1985. This hydropower plant consists of a power plant, water lock, spillway and non-spillway dam, as well as a dam crossing in the middle of which there is a border between two countries.

The spillway dam consists of seven spillways, each 21 m wide. Seven pendulums were installed to measure radial and tangential deformations. In this paper, the tangential displacement of the points in the first spillway and fourth spillway of the dam is analysed with the proposed method. These points are denoted by F1 and F4. The data set included 729 data samples. They were divided into training and test sets. The data from January 1997 to December 2007 were used to train, and the data from January 2008 to December 2010 were used to test, Fig. 1.

3. NARX system identification

An important and useful class for simulation of dam structural behaviour is the nonlinear autoregressive with exogenous inputs model [14,31]. The nonlinear model for prediction of the dam tangential displacement has two inputs (u_1, u_2) and one output (z_m) and can be described as follows:

$$z_m(k) = f_m(\boldsymbol{\varphi}(k), \boldsymbol{\theta}) \tag{1}$$

where $z_m(k)$ is the output of the model, k is the time instant, f_m is the unknown nonlinear function, $\boldsymbol{\varphi}(k) = (z(k-1), z(k-2), \ldots, z(k-n_z), u_1\ (k-1), u_1(k-2), \ldots, u_1(k-n_{u1}), u_2(k-1), u_2(k-2), \ldots, u_2(k-n_{u2}))$ is the regression vector, $\boldsymbol{\theta}$ is the parameter vector, n_{u_1} and n_{u_2} denote the numbers of the lags of the inputs (u_1, u_2) and n_z denotes the number of the lags of the output (z).

The problem with identification of the nonlinear structural behaviour is to approximate the unknown function f_m in (1) from the sampled data $\{(u_1(k),\ u_2(k), z(k))|k = 1, 2, \ldots, p\}$, where p is the number of the sample data. Identification of nonlinear structural behaviour is a difficult task because the nonlinear function can be assumed in different forms. In this paper, support vector regression is used for approximation of the unknown nonlinear function.

The selection of an appropriate set of input variables and the regressors defined by the input and output lags is important for obtaining a high-quality model. Many of the described methods for regressor selection are based on heuristics, expert knowledge, statistical analysis [32], or a combination of these.

The input vector to the SVR model consists of z and u_1 and u_2 which are past values of the output and input, respectively:

$$\boldsymbol{x}^T = [z(k-1), z(k-2), \ldots, z(k-n_z), u_1(k-1), u_1(k-2), \ldots, u_1(k-n_{u1}), u_2(k-1), u_2(k-2), \ldots, u_2(k-n_{u2})] \tag{2}$$

The output is $y = z_m(k)$.

4. Support vector regression

Consider a training data set $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_p, y_p)\} \in \mathbb{R}^N \times \mathbb{R}$, where x_i is a vector of input variables and y_i is the corresponding output value, p is the number of training data points. In ε-SV regression [33], the goal is to find a function $f(\boldsymbol{x})$ that has the most ε deviation from the actually obtained targets y_i for all the training data.

The SVM approximates the function in the following form:

$$f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{\phi}(\boldsymbol{x}) \rangle + b \text{ with } \boldsymbol{w} \in \mathbb{R}^N, \quad b \in \mathbb{R} \tag{3}$$

where $\boldsymbol{w}$ denotes the weight vector, $\boldsymbol{\phi}(\boldsymbol{x})$ represents the high-dimensional feature spaces which is nonlinearly transformed from the input space $\boldsymbol{x}$, $\langle \cdot,\cdot \rangle$ denotes the dot product between $\boldsymbol{w}$ and $\boldsymbol{\phi}(\boldsymbol{x})$ and b is a bias.

The weight vector ($\boldsymbol{w}$) and the bias (b) can be estimated by minimizing the regularized risk function:

$$R(C) = \frac{1}{2}|\boldsymbol{w}|^2 + C\frac{1}{p}\sum_{i=1}^{p} L_\varepsilon(f(\boldsymbol{x}_i), y_i) \tag{4}$$

In the regularized risk function, the first term $\frac{1}{2}|\boldsymbol{w}|^2$ is the regularized term that controls the function capacity. The second term $\frac{1}{p}\sum_{i=1}^{p} L_\varepsilon(f(x_i), y_i)$ is the empirical error. C is the regularization constant that determines the trade-off between the empirical risk and the regularization term. In the SVM regression, ε-insensitive loss function is most commonly used:

$$L_\varepsilon(f(\boldsymbol{x}_i), y_i) = \begin{cases} |f(\boldsymbol{x}_i) - y_i| - \varepsilon, & |f(\boldsymbol{x}_i) - y_i| \geqslant \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where ε is a constant called the tube size.

Minimization of (4) is equivalent to solving the following primal optimization problem:

$$\text{minimize} \frac{1}{2}|\boldsymbol{w}|^2 + C\frac{1}{p}\sum_{i=1}^{p}(\xi_i + \xi_i^*) \tag{6}$$

subject to,

$$\begin{cases} y_i - \langle \boldsymbol{w}, \boldsymbol{\phi}(\boldsymbol{x}_i) \rangle - b \leqslant \varepsilon + \xi_i, & i = 1, 2, \ldots, p \\ \langle \boldsymbol{w}, \boldsymbol{\phi}(\boldsymbol{x}_i) \rangle + b - y_i \leqslant \varepsilon + \xi_i^*, & i = 1, 2, \ldots, p \\ \xi_i, \xi_i^* \geqslant 0, & i = 1, 2, \ldots, p \end{cases}$$

Positive slack variables ξ_i and ξ_i^* represent the distance from the actual values to the corresponding boundary values of the ε tube. A schematic representation of the SVR using the ε-insensitive loss function is shown in Fig. 2.