Expert Systems with Applications 36 (2009) 7624–7629

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Generalization performance of support vector machines and neural networks in runoff modeling

Mohsen Behzad [a,*], Keyvan Asghari [a], Morteza Eazi [a], Maziar Palhang [b]

[a] Isfahan University of Technology, Civil Engineering Department, 8415683111, Isfahan, Iran
[b] Isfahan University of Technology, Electrical and Computer Department, Iran

ARTICLE INFO

Keywords:
Support vector machines (SVMs)
Artificial neural networks (ANNs)
Runoff prediction
Rainfall and climate data

ABSTRACT

Effective one-day lead runoff prediction is one of the significant aspects of successful water resources management in arid region. For instance, reservoir and hydropower systems call for real-time or on-line site-specific forecasting of the runoff. In this research, we present a new data-driven model called support vector machines (SVMs) based on structural risk minimization principle, which minimizes a bound on a generalized risk (error), as opposed to the empirical risk minimization principle exploited by conventional regression techniques (e.g. ANNs). Thus, this stat-of-the-art methodology for prediction combines excellent generalization property and sparse representation that lead SVMs to be a very promising forecasting method. Further, SVM makes use of a convex quadratic optimization problem; hence, the solution is always unique and globally optimal. To demonstrate the aforementioned forecasting capability of SVM, one-day lead stream flow of Bakhtiyari River in Iran was predicted using the local climate and rainfall data. Moreover, the results were compared with those of ANN and ANN integrated with genetic algorithms (ANN-GA) models. The improvements in root mean squared error (RMSE) and squared correlation coefficient (R^2) by SVM over both ANN models indicate that the prediction accuracy of SVM is at least as good as that of those models, yet in some cases actually better, as well as forecasting of high-value discharges.

1. Introduction

River flow forecasting is required to provide basic information on a wide range of problems related to the design and operation of river systems. The availability of extended record of rainfall and other climate data, which could be used to obtain stream flow data, initiated the practice of rainfall-runoff modeling.

Conceptual or physically-based models are of great importance in the understanding of hydrological processes. But these modeling approaches are limited due to multitude as well as complexity of the processes involved and also by scarcity of data (Task Committee on Application of the Artificial Neural Networks in Hydrology, 2000a). In recent years, non-linear data-driven models are being widely used as surrogate for the conceptual models. They are able to capture the behavior of the underlying physical or other processes. Such approaches might be made to evolve reliable forecasting models using measured historical data. These modeling tools like artificial neural networks (ANNs) do not require knowledge of mathematical relationship between the inputs and corresponding outputs as well as explicit characterization and quantification of physical properties and conditions.

In addition to applications of ANNs, in the past decade support vector machines (SVMs) have gained the attention of many researchers. SVMs rely on statistical learning theory which enables learning machines to generalize well to unseen data. The main characteristics of the SVM can be summarized as follows: (a) a global optimal solution is found by quadratic programming method, (b) the result is a general solution which relatively avoids over training as it employs the structural risk minimization principle, (c) the solution is also sparse, hence only a limited subset of training points contribute to this solution, and (d) non-linear relations can be learned effectively owing to the usage of kernel functions. Consequently, SVMs have gained popularity in many traditionally ANNs dominated applications. They seem to be powerful alternatives which overcome some of the basic weakness related to ANNs modeling while retain all strengths of ANNs (Task Committee on Application of the Artificial Neural Networks in Hydrology, 2000b). The objective of this study is to provide a forecasting system based on SVM approach whose ability is to predict one-day lead stream flow by emphasizing high-value discharges. The results associated with ANN models are used to demonstrate the performance of SVM method.

* Corresponding author.
E-mail address: m_behzad@cv.iut.ac.ir (M. Behzad).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.09.053

Although SVMs have remarkable successes in various fields, there are a few studies on their applications in water resources and hydrology (Smola & Schölkopf, 2004). Dibike, Velickov, Solomatine, and Abbott (2001) demonstrated the capability of SVM in hydrological prediction when it was applied in classifying digital remote sensing data and also modeling of rainfall-runoff - process. They also compared the results with those of ANN model (Dibike et al., 2001). Liong and Sivapragasam (2002) successfully employed SVM in the flood stage forecasting (Liong & Sivapragasam, 2002). Asefa, Kemblowski, Urroz, McKee, and Khalil (2004), Asefa, Kemblowski, Urroz, and McKee (2005a) examined the capability of SVM in optimal design of groundwater monitoring networks (both for head observation and contamination monitoring networks) and inferred that the SVM can be used as a novel tool for selecting the configuration of optimal monitoring network (Asefa, Kemblowski, Urroz, & McKee, 2005a; Asefa, Kemblowski, Urroz, McKee, & Khalil, 2004). Moreover, they investigated some other successful applications of SVM corresponding to hydrology and water resources management such as snow-runoff modeling and learning of chaotic time series (Asefa, Kemblowski, Lall, & Urroz, 2005b; Asefa, Kemblowski, McKee, & Khalil, 2006). Khalil, Almasri, McKee, and Kaluarachchi (2005a) examined four kinds of learning algorithms named SVMs, RVMs (relevance vector machines), LWPR (locally weighted projection regression) and ANNs as surrogates for relatively complex and time consuming mathematical models, to estimate the groundwater nitrate contamination level (Khalil et al., 2005a). He also applied RVM, which is a probabilistic model, with SVM in capturing the uncertainties in both model parameters and input data with an application in a real case study involving the real-time operation of a reservoir in a watershed in southern Utah (Khalil, McKee, Kemblowski, & Asefa, 2005b). Khalil, McKee, Kemblowski, Asefa, and Bastidas (2006) exploited the appealing regularization concepts of both SVM and RVM to determine the optimal parameters of powerful state-space reconstruction methodology and also hyper parameters of learning machines within a multi objective optimization framework and finally analyzed chaotic dynamic systems (Khalil et al., 2006). Yu, Chen, and Chang (2006) invoked SVM to establish a river stage forecasting model, whose input vector encompassed both rainfall and river stage, to predict the hourly stage of the flash flood (Yu et al., 2006). Lin, Cheng, and Chau (2006) used SVM for long-term discharge prediction and compared the performance of SVM with two alternative methods, ANN and ARMA (Auto Regression Moving Average) models (Lin et al., 2006). In all of the above-mentioned applications, when researchers collated the results of SVM modeling with an alternative modeling approach such as ANN, it would be obvious that SVM had promising performance due to its high generalization characteristic. It should be noted that the use of ANNs in prediction of runoff has been investigated by many researchers (e.g. Dibike & Solomatine, 2001; Kisi, 2004; Tokar & Markus, 2000; Zealand & Burn, 1999).

The rest of this paper is organized as follows. The SVR model is introduced in the next section and the third section illustrates the study area and available data. Then, support vectors machines and neural networks characteristics applied in this study are presented in the fourth section. Subsequently, the results of the SVMs along with ANNs modeling are depicted in the fifth section to demonstrate the performances of different forecasting models. Some conclusions are then made in the final section.

2. SVMs theory

SVMs are particular learning systems that use a linear high dimensional hypothesis space called feature space. These systems are trained using a learning algorithm which is based on optimization theory. This method was introduced by Vapnik (1998) and his colleagues as a robust and significant learning tool, which uses a learning bias derived from statistical learning theory (SLT). We borrow the following SVMs theoretical background from Schölkopf and Smola (2002), Cristianini and Shawe-Taylor (2000).

2.1. SVMs for regression estimation

SVMs have been employed for regression estimation, the so called support vector regression (SVR), in which the real value functions are estimated. In this case, the aim of learning process is to find a function $f(\mathbf{x})$ as an approximation of the value $y(\mathbf{x})$ with minimum risk, and only based on the available independent and identically distributed data, i.e.

$$(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m) \subseteq (\mathbf{X} \subseteq \mathbf{R}^{\mathbf{n}} \times Y \subseteq \mathbf{R}) \tag{1}$$

In SVR algorithm, the estimation function is determined by a small subset of training samples namely support vectors (SVs). Also in this algorithm, a specific loss function called ε-insensitive loss is developed to create a sparseness property for SVR. This function is described as follows:

$$|y - f(\mathbf{x})|_\varepsilon = \begin{cases} 0 & \text{if } |y - f(\mathbf{x})| \leqslant \varepsilon \\ |y - f(\mathbf{x})| - \varepsilon & \text{otherwise} \end{cases} \tag{2}$$

where $f(\mathbf{x})$, which is computed by the SVR, is the estimated value of the y and the corresponding errors being less than ε*boundary* (ε*-tube*) are not penalized (Fig. 1).

For developing the regression algorithm, we begin with the linear function estimation. It is clear that every linear function of input vector $\mathbf{x}$ has the following representation:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \quad \text{where } \mathbf{w}, \mathbf{x} \in X \subseteq R^n, b \in \mathbf{R}. \tag{3}$$

Noted that angle bracket indicates the inner product of two vectors in Hilbert space (i.e. a space in which inner product of two vectors has a real value, also called inner (or dot) product space).

For finding $f(\mathbf{x})$ one should minimize the regulated risk functional (R_{reg}) (instead of just empirical risk functional which is used in traditional learning algorithms such as ANNs) defined as follows;

$$R_{\mathrm{reg}}[f] = \frac{1}{2}\|\mathbf{w}\|^2 + C.R^{\varepsilon}_{\mathrm{emp}}[f] \quad \text{where } R^{\varepsilon}_{\mathrm{emp}}[f] = \frac{1}{m}\sum_{i=1}^{m} |y_i - f(\mathbf{x}_i)|_\varepsilon \tag{4}$$

The $R^{\varepsilon}_{\mathrm{emp}}$ is the empirical error over training data which is defined in ε-insensitive loss function framework. Coefficient C in the Eq. (4) is an indicator of the complexity of function f. Briefly speaking, the minimization of the R_{reg} illustrates the principle idea of the structural risk minimization theory which states that for achieving the minimum risk, simultaneous control of the complexity of the model and the error owing to training data is essential. This idea improves the generalization of the SVRs.

2.2. Principle objective function

It has been shown that minimizing the Eq. (4) is equivalent to the following convex constrained quadratic optimization problem:

$$\begin{aligned} \text{minimize} \quad & \tau(\mathbf{w}, \xi^{(*)}) = \tfrac{1}{2}\|\mathbf{w}\|^2 + C \cdot \tfrac{1}{m}\sum_{i=1}^{m}(\xi_i + \xi_i^*), \\ \text{subject to} \quad & (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - y_i \leqslant \varepsilon + \xi_i, \\ & y_i - (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \leqslant \varepsilon + \xi_i^*, \\ & \xi_i^{(*)} \geqslant 0. \end{aligned} \tag{5}$$

$\mathbf{w} \in \mathbf{X}, \xi_{\mathbf{i}}^{(*)} \in \mathbf{R^m}, b \in \mathbf{R}, \quad i = 1, \dots, m$

where the slack variable $\xi^{(*)}$ encompasses both the ξ, ξ^* variables.