Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 186 (2017) 537 – 543

Procedia Engineering

www.elsevier.com/locate/procedia

XVIII International Conference on Water Distribution Systems Analysis, WDSA2016

Support vector machines in urban water demand forecasting using phase space reconstruction

Sina Shabani[a], Peyman Yousefi[a], and Gholamreza Naser[a]*

[a]*Okanagan School of Engineering, The University of British Columbia, Canada*

Abstract

High complexity of water distribution systems' (WDS) dynamics has convinced researchers to look for more sophisticated statistical approaches in urban water demand forecasting. Given the huge threat to major water reserves and ongoing draughts, water authorities are concerned with long term analysis of water demand to deal with uncertain future of this dynamic infrastructure. Researchers have tried a wide range of modelling techniques to propose an accurate model. However, applications of machine learning techniques are yet to be explored in detail. This research proposes a support vector machine (SVM) model, using polynomial kernel function to forecast monthly water demand of City of Kelowna (CKD), Canada. The prime objective of this research is to assess the use of phase space reconstruction prior to design of models' input variables combinations. Results of this study proved optimum lag time of the input variables can significantly improve the performance of SVM models.

Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems

Keywords: Water demand forecasting; phase space reconstruction; average mutual information; lag time; support vector machine.

1. Introduction

Scientific and technological awareness of engineers have been constantly improving both practically and theoretically. Like other engineering disciplines, water resources engineering is coupled with new scientific concepts using data mining, management, computer programs, and artificial intelligence. Potable water reserves have been under a serious threat due to water scarcity concerning many countries in the world. Therefore, forecasting

* Corresponding author. Tel.: +1-250-807-8464; fax: +1-250-807-9850.
E-mail address: bahman.naser@ubc.ca

1877-7058
Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems
doi:10.1016/j.proeng.2017.03.267

municipal water demand is essential to all water utilities for water distribution systems (WDS) planning, design, operation, and asset management. With the current water scarcity, robust and accurate demand forecasting models are needed to aid WDS operators with meaningful data for better judgment calls in their design and operation. Accurate predictions of water demand can enhance water sectors with tools for efficient management of reservoirs and water resources, supply of water, and design of hydraulic structures. Researchers have proposed different techniques for predicting this parameter. Given the complexity in water demand analysis and existence of a large set of effective determinants, researchers have employed intelligent techniques for more accurate models. These emerging artificial intelligent techniques are receiving more attention than conventional parametric and stochastic models. Conventional methods of demand forecasting have been time series prediction and linear regression [1-3] which do not account for nonlinearity of the input data in such models. Therefore, some new techniques such as ANN (artificial neural networks) [4-6] and Non-linear regression [7] have been recently proposed which could outperform traditional linear models. Support vector machine method was firstly proposed as a classification method in 1995 [8]. However, it was later modified as a regression technique. Using kernel functions, SVM regressions can account for nonlinearity in the systems. SVMs have been recently used in predictive models [9]; however, handful number of studies deployed this technique in water demand forecasting [10, 11]. Therefore, there is a need for further investigation of these models. For the first time, this research applies SVMs in urban water demand forecasting along with pre-processed water demand and climatic information. Time series of input data are transformed using phase space reconstruction, which allows different lag times to be explored in the final output of the model.

2. Methodology

The main purpose of this approach was to determine the impact of lag time on support vector machines using phase space reconstruction. AMI (average mutual information) is opted for determination of optimum lag time. This technique is selected rather than auto-correlation function (ACF) and correlation integral (CI) as these alternatives require large sets of data or fail to exhibit nonlinearity of the models [12].

2.1. Phase Space Reconstruction: proper lag time

The following equation (1) can determine the appropriate lag time between two independent time series. This approach uses the joint probability of sequential time series which succeed one another by an increment (equal to 1unit of the lag time). Moreover, the product of their marginal probability is also utilized to determine the optimum lag time. Similar to Shannon's entropy, this technique can be a good estimate of how entropy levels can change the dynamics of these deployed time series in forecasting models. An optimum lag time is used to make sure sufficient information is added on the balanced independent time series which can magnify the behavior of these time series in a desired phase space. Figure 1 shows the first local minimum of the drawn graphs can be a fair estimate of optimum lag time. In this case, optimum lag time was selected as 1 month for precipitation. On the other hand, temperature and water demand showed an optimum lag time of 3 months.

$$I_\tau = \sum_{i=1}^{i=n} P(X_i, X_{i+\tau}) \cdot \log_2 \frac{P(X_i, X_{i+\tau})}{P(X_i) \cdot P(X_{i+\tau})} \tag{1}$$