Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 119 (2015) 844 – 853
Procedia Engineering
www.elsevier.com/locate/procedia

13th Computer Control for Water Industry Conference, CCWI 2015

Short-term forecasting of hourly water consumption by using automatic metering readers data

Antonio Candelieri[a,b] *, Davide Soldi[a], Francesco Archetti[a,b]

[a]Consorzio Milano Ricerche, via Roberto Cozzi 53, Milano, 20125, Italy
[b]Department of Computer Science, Systems and Communication, University of Milano-Bicocca, viale Sarca 336, Milano, 20126, Italy

Abstract

A completely data-driven, fully adaptive self-learning algorithm for water demand forecasting in the short-term and with hourly periodicity is proposed, according to the renewed interest generated by the availability of new technological solutions such as Automatic Metering Readers (AMR), a key enabler of the "Smart Water" paradigm. The approach is based on two sequential stages: at the first stage (time-series clustering) the daily water demand patterns (i.e., time-series of hourly data) are analysed to identify a limited set of typical behaviours. At the second stage Support Vector Machine regression is used to obtain one specific forecasting model (consisting of a regression model for each hour) for each cluster identified at the first stage. The approach has been validated on real data acquired by AMRs deployed on the Italian pilot site of ICeWater, computing the widely adopted error measure MAPE (Mean Absolute Percentage Error).

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: Short-term demand forecasting; Time-series mining; Time-series clustering; Support Vector Machines regression

1. Introduction

Water Distribution Networks (WDN) are large-scale systems characterized by complex decision making activities. WDM managers need to reliably estimate the water demand in the short-term (typically 1 day ahead), in order to operate their reservoirs and treatment plants appropriately to meet demand while energy-related costs for caption, treatment and pumping. In the United States, examples of the adoption of a control model based on a short-

* Corresponding author. Tel.: +39-02-6448-2185.
E-mail address: candelieri@milanoricerche.it

Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.948

term water demand forecasts were validated already in 2009. [1]. In the Netherlands the penetration of short-term forecasting models is expected to rise over 90% in 2016 [2].

Supervisory Control And Data Acquisition (SCADA) systems are currently widely adopted by water utilities, usually providing time-series data about overall system or region-wide demand, typically with a resolution of 10-15 minutes. Nowadays, the availability of smart metering devices, such as Automatic Metering Readers (AMRs), makes possible the application of water demand forecasting also at individual users/meters level, even if with a lower resolution (usually 1 hour or, in the best case, 30 minutes). These technological solutions have generated a renewed interest in a new generation of individual demand forecasting as enabler of the “Smart Water” paradigm.

1.1. State of the art of water demand short-term forecasting

A recent review about urban water demand forecasting is provided in [3]: a wide variety of approaches has been reviewed, their application differs according to the management objectives as well as the variable to be forecasted, the *forecast periodicity* and the *forecast horizon*. Furthermore, the availability and choice of specific determinants can influence the selection of the forecasting approach to use. However, methods and models whose input variables can be easily collected, monitored and used by the utility (i.e. consumption data) should be preferred for practical application, reducing the risk to add noise/errors coming from data/information sources which are not under control (e.g. weather forecast provided by external sources/services).

A possible classification of the current water demand short-term forecasting approaches consists in distinguishing between *linear* and *nonlinear* methods [4]. Usually linear methods are not so effective due to the intrinsic nonlinearity into water demand data. Indeed, as reported in [5], most of the existing demand forecasting models can be divided into two groups: those modelling the time series behaviour and those predicting it. One is devoted to model specific components such as periodicity (seasonality) and trends, the other one is, usually, an autoregressive model using short memory data and reproducing the underlying “generation” process of data. The main disadvantage is the limited predictability of water demand at sub-daily scale, due to the nonlinearities of the problem. The most relevant differentiation proposed in [3] refers to the inclusion/exclusion of exogenous variables to build the water demand forecasting model, concluding with some more general approaches, in particular Artificial Neural Networks (ANNs), which can be – and have been – applied in both the two mentioned cases. ANN is a machine learning approach widely applied to forecast water demand. Research papers about ANN for water demand forecasting typically involve a comparison of the performance between different ANN models and more conventional regression models [6][7][8][9] as well as (univariate) time series models [10]. Furthermore, hybrid approaches have been also proposed, in [11] linear regression is used to model the deterministic component of water demand and Artificial Neural Networks (ANNs) are used to model the cyclical component. As result, the composite model offers more accurate forecasts with respect to those obtained from linear regression and ANN separately.

Recently some advances have been achieved in the application of machine learning techniques for water demand forecast. In particular not only ANNs have been adopted in the last years, but also more effective and efficient strategies such as Support Vector Machine (SVM) regression [5][12][13][14]. Analogously, meta-heuristic based approaches gained a renewed interest from their application to optimize the parameters setting for a specific machine learning algorithm, such as evolution-based strategies (e.g., Genetic Programming, Genetic Algorithms, etc.). A relevant example is the approach proposed in [4] which uses Evolutionary Artificial Neural Networks (EANN). Recently a heuristic known as Teaching-Learning-based Optimization (TLBO), emulating the effect of a teachers to learners in a class [15], has been improved (Ameliorated TLBO, ATLBO) and used to optimally configure the parameters of a Least Square Support Vector Machine (LS-SVM), that is a an extension of SVM usually preferred in the case of large scale problem [15].

With respect to SVM regression, it proved to be the best computational model for forecasting hourly water demand when compared with other different approaches, such as ANNs, Projection Pursuit Regression (PPR), Multivariate Adaptive Regression Splines (MARS), Random Forests and weighted pattern-based water demand forecasting [9]. More recently, Multiple Kernel Learning has been proposed in order to improve accuracy of SVM for water demand forecasting. In [12] a Multiple Kernel regression (MKr) has been proposed to extend SVM regression through a combination of different kernels from as many types as kinds of input data source are available. Moreover, the paper focuses on water demand forecasting in the presence of a continuous source of information,