Discovering Weekly Seasonality for Water Demand Prediction using Evolutionary Algorithms

Piotr Lipinski
Patryk Filipiak
Pawel Rychlikowski
Computational Intelligence Research Group, Institute of Computer Science, University of Wroclaw, Poland
piotr.lipinski@cs.uni.wroc.pl

Justyna Stanczyk
Joanna Kajewska-Szkudlarek
Janusz Lomotowski
Institute of Environmental Engineering, Wroclaw University of Environmental and Life Sciences, Poland

Tomasz Konieczny
New Technologies Centre, MPWiK S.A., Wroclaw, Poland

ACM Reference format:
Piotr Lipinski, Patryk Filipiak, Pawel Rychlikowski, Justyna Stanczyk, Joanna Kajewska-Szkudlarek, Janusz Lomotowski, and Tomasz Konieczny. 2017. Discovering Weekly Seasonality for Water Demand Prediction using Evolutionary Algorithms. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 2 pages.
DOI: http://dx.doi.org/10.1145/3067695.3082049

1 INTRODUCTION

The modern approach to water supply network management and operation is related to the use of modern solutions from both technical and strategic perspectives. Apart from practices promoted by International Water Association (IWA) (active leakage control, pressure management, speed and quality of repair, pipeline and assets management), water demand prediction systems are the future. Preparing short-, medium- and long-term water consumption forecasts is the key factor these days. Short-term simulations, mainly those covering the period of 24 or 48 hours, are used to optimise the operation of pumping stations and to resolve current exploitation issues, whereas long-term analyses, covering more than one month or year, are said to support the decision-making process regarding the design and development of water supply networks. Medium-term predictions, covering weeks, are used to create time schedules for the maintenance of water supply networks and develop failure prevention procedures. Due to the lack of assessment teams and proper assessment tools, the majority of water and sewerage companies store registered time series without thorough study of the data. Information regarding the current operating status of the water supply network included in the series is not properly used and irretrievably lost.

Owing to the fast growth of IT, new algorithms for multicriterial optimisation of water supply network issues can be implemented. Until now, water demand predictions were made, depending on the time frame, with the use of, among others, Support Vector Machine (SVM) [1], Evolutionary Artificial Neural Networks (EANNs) [4], a hybrid of Support Vector Regression and Adaptive Fourier Series (SVR-AFS) [2] and Dynamic Artificial Neural Network (DANN) [3].

GECCO '17 Companion, Berlin, Germany
 978-1-4503-4939-0/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3067695.3082049

The quality of drinking water and its quantity are monitored. The quantity control is performed for water balance purposes. To validate hydraulic phenomena inside water pipes the following parameters are usually measured: pressure, flow rate and volume. Measuring instruments are fixed to pipes carrying water to District Metering Areas in water metre wells. These instruments are magnetic flow meters and pressure transmitters/manometers. Time series registered with a time-step of 10-15 minutes, are parts of a reporting and network control system. A GSM module is used to transmit data to the water supply network operator in real time. The water supply network of Wroclaw, Poland consists of almost 30 District Metering Areas (DMAs) equipped with more than 80 measurement points. Additionally, the monitoring system uses meteorological stations. They gather weather information including precipitation level, air temperature, relative air humidity, wind speed and direction, sunshine duration and air pressure. One-minute time-steps are usually used for measuring. The data is sent to a datalogger via a wireless GSM/GPRS connection.

2 PROBLEM DEFINITION

Our research focuses on water demand prediction based on the data from the water supply network and the meteorological stations (real-world data from Wroclaw in Poland) by computational intelligence methods. This abstract describes an approach to improve the prediction results by discovering weekly seasonality in water demand time series using evolutionary algorithms.

The water demand is predicted using Support Vector Regression (SVR) with Radial Basis Functions (RBF) and the following parameter settings: $C = 1.0$ and $\gamma = 0.125$, described in details in [5]. The input data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ consist of the input vectors $\mathbf{x}_t \in \mathbb{R}^8$, for each day $t = 1, 2, \ldots, T$ of a given period, that contain the meteorological data (air temperature, atmospheric pressure, relative humidity, sunshine duration), the daylength and the water demand observed in the previous day. The output data $\mathrm{Y} = \{y_1, y_2, \ldots, y_T\}$ (to be predicted by SVR) consist of the corresponding output values $y_t \in \mathbb{R}$ that describe the observed water demand. In both, input and output data, the water demand comes from the time series Q_t preprocessed earlier with removing trends and seasonalities in such a way that $Q_t = Q_t^{(O)}/Q_t^{(T)} - Q_t^{(S)}$, where $Q_t^{(O)}$ denotes the original time series of the observed water demand, $Q_t^{(T)}$ denotes the trend defined by the 7-days moving average and $Q_t^{(S)}$ denotes the weekly seasonality. SVR is constructed on a train dataset concerning a train period, from September 1, 2014 to February 29, 2016, and tested on a test dataset concerning a test period, from March 1, 2016 to November 30, 2016.

Removing the seasonalities and trends from data are difficult issues in water demand prediction. Note that such time series are strongly correlated with weather conditions. There are also some irregularities related to movable feasts, renovations or break-downs. Moreover, some trends are present due to the urban expansion.

3 EVOLUTIONARY APPROACH

Instead of time series modeling methods, which usually require long and stable historical time series to efficiently determine the seasonality, we use an evolutionary approach to discover the weekly seasonality $Q_t^{(S)}$. The search space $\Omega = \mathbb{R}^7$ is composed of 7-dimensional vectors $\mathbf{s} = (s_1, s_2, \ldots, s_7)$ defining the weekly seasonality in such a way that s_i denotes the ratio of the total weekly water demand in the i-th day of the week. The objective function $F : \Omega \rightarrow \mathbb{R}$ is the Mean Average Percentage Error (MAPE) [5] of SVR constructed on the train dataset preprocessed with the weekly seasonality model defined by the candidate solution. In order to solve such an optimization problem, the Evolution Strategy algorithm [6] is used with a population of 200 candidate solutions evolved over 50 iterations, with Global Intermediate Recombination [6] and with mutation parameters $\boldsymbol{\sigma} \in \mathbb{R}^7$ encoded in each candidate solution and managed by the autoadaptation mechanism as in the regular ES(μ, λ) algorithm, described in details in [6]. It is worth noticing that evaluation of each candidate solution requires training SVR each time anew, which makes solving the optimization problem time-consuming.

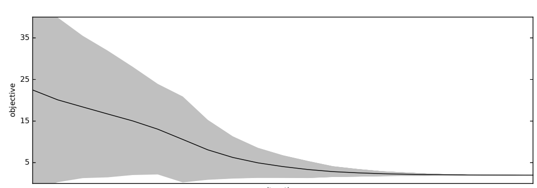

Figure 1: The values of the objective function in successive iterations of ES minimizing MAPE of SVR on the train dataset (September 1, 2014 – February 29, 2016)

Figure 1 presents the values of the objective function in successive iterations of ES minimizing MAPE of SVR on the train dataset. The black line presents the mean value of the objective function evaluated on the 200 candidate solutions from the current population, and the gray area corresponds to the standard deviation. The decreasing curve plotted in Figure 1 shows that our evolutionary algorithm was able to minimize the objective function F, and thus provide an efficient weekly seasonality model.

4 RESULTS

The experiments concern the water demand prediction in the water supply network of Wroclaw, Poland and is based on real-world data coming from MPWiK S.A., Wroclaw, Poland. The baseline was defined by determining the weekly seasonality by the regular time series modeling methods that led to MAPE on the train dataset equal to 2.0080% and on the test dataset equal to 3.2233%. SVR constructed on the train dataset preprocessed with the weekly seasonality model defined by the best candidate solution found by the evolutionary algorithm led to better results with MAPE on the train dataset equal to 1.9826% and on the test dataset equal to 3.1714%. It is worth noticing that the evolutionary approach did not use any calendar knowledge concerning the dates of the time series.

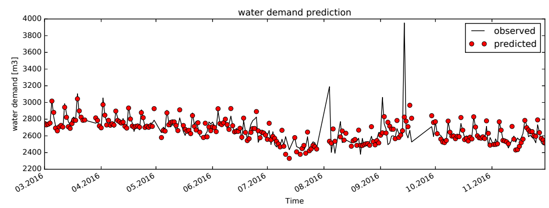

Figure 2: Water demand prediction on the test dataset (March 1, 2016 - November 30, 2016): observed and predicted values

Figure 2 presents the results of the prediction on the test dataset. Note that the predicted water demands correspond to the observed values, except few cases with sudden increase of water consumption, caused by some special events, such as World Youth Days in Wroclaw (impossible to predict without the extra knowledge). Figure 3 presents the differences between the observed and the predicted values of water demand.

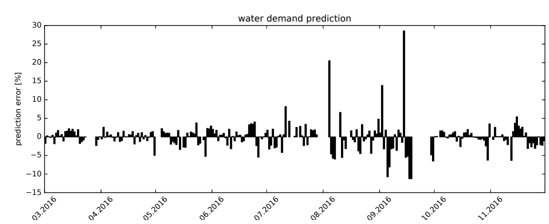

Figure 3: Water demand prediction on the test dataset (March 1, 2016 - November 30, 2016): prediction errors

REFERENCES

[1] A. Candelieri, F. Archetti, Identifying Typical Urban Water Demand Patterns for a Reliable Short-Term Forecasting - The Icewater Project Approach, Procedia Engineering - 16th Conference on Water Distribution System Analysis, WDSA 2014, no. 89, pp. 1004-1012, 2014.
[2] B. M. Brentan, E. Luvizotto Jr., M. Herrera, J. Izquierdo, R. Perez-Garcia, Hybrid regression model for near real-time urban water demand forecasting, Journal of Computational and Applied Mathematics, no. 309, pp. 532-541, 2017.
[3] M. Ghiassi, F. Fa'al, A. Abrishamchi, Large metropolitan water demand forecasting using DAN2, FTDNN, and KNN models: A case study of the city of Tehran, Iran, Urban Water Journal, vol. 0, pp. 1-5, 2016.
[4] R. Romano, Z. Kapelan, Adaptive water demand forecasting for near real-time management of smart water distribution systems, Environmental Modelling & Software, no. 60, pp. 265-276, 2014.
[5] A. Smola, B. Scholkopf, A tutorial on support vector regression, Statistics and Computing, vol. 14, no. 3, pp. 199-222, 2004.
[6] H.-P., Schwefel, Evolution and Optimum Seeking: The Sixth Generation, John Wiley & Sons, 1993.