ARTICLE IN PRESS

Environmental Modelling & Software xxx (2017) 1-15



Contents lists available at ScienceDirect

Environmental Modelling & Software



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

Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques

Ll. Corominas ^{a, *}, M. Garrido-Baserba ^b, K. Villez ^c, G. Olsson ^d, U. Cortés ^e, M. Poch ^f

^a ICRA, Catalan Institute for Water Research, Scientific and Technological Park of the University of Girona, H₂O Building, Emili Grahit 101, 17003 Girona, Catalonia, Spain

^b Department of Civil and Environmental Engineering, University of California, Irvine, CA 92697-2175, USA

^c Department of Process Engineering, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, PO Box 611, 8600 Dübendorf, Switzerland

^d Industrial Electrical Engineering and Automation, Lund University, Box 118, SE-22100 Lund, Sweden

^e KEMLg, Universitat Politècnica de Catalunya, Barcelona Supercomputing Center, Edifici Omega 205d, Barcelona 08034, Catalonia, Spain

^f LEQUIA, Institute of the Environment, University of Girona, Campus Montilivi, 17071 Girona, Catalonia, Spain

ARTICLE INFO

Article history: Received 23 February 2017 Received in revised form 4 October 2017 Accepted 9 November 2017 Available online xxx

Keywords: Data mining Data processing Data quality WWTP Knowledge

ABSTRACT

The aim of this paper is to describe the state-of-the art computer-based techniques for data analysis to improve operation of wastewater treatment plants. A comprehensive review of peer-reviewed papers shows that European researchers have led academic computer-based method development during the last two decades. The most cited techniques are artificial neural networks, principal component analysis, fuzzy logic, clustering, independent component analysis and partial least squares regression. Even though there has been progress on techniques related to the development of environmental decision support systems, knowledge discovery and management, the research sector is still far from delivering systems that smoothly integrate several types of knowledge and different methods of reasoning. Several limitations that currently prevent the application of computer-based techniques in practice are highlighted.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

How do we turn passive data into actionable knowledge or something compelling that improves wastewater treatment operation or supports decision-making? The aim of this paper is to describe the state-of-the art computer-based techniques for data analysis as applied in the context of wastewater treatment

Corresponding author.

E-mail addresses: lcorominas@icra.cat (Ll. Corominas), manelg@uci.edu (M. Garrido-Baserba), Kris.Villez@eawag.ch (K. Villez), Gustaf.Olsson@iea.lth.se (G. Olsson), ia@cs.upc.edu (U. Cortés), manuel.poch@udg.edu (M. Poch).

operation. This critical review targets method developers (mostly within the research community) by discussing the evolution of a selection of methods and identifying limitations of method development and selection, as well as plant managers and software developers by identifying barriers that limit bringing methods into practice.

This paper is structured as follows: first, we briefly define the driving forces within the wastewater treatment field that pushed for the development of computer-based techniques for data analysis. Second, we describe the variety of available techniques that enable the transformation of data into information and, beyond that, into knowledge by means of a review of the techniques applied thus far in wastewater treatment plants (WWTPs). In parallel, a critical analysis of the maturity and temporal evolution of each technique is given. Finally, a discussion is provided on the limitations in this field.

1.1. Driving forces

WWTPs treat wastewater collected from households and

https://doi.org/10.1016/j.envsoft.2017.11.023 1364-8152/© 2017 Elsevier Ltd. All rights reserved.

Please cite this article in press as: Corominas, Ll.-g., et al., Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques, Environmental Modelling & Software (2017), https://doi.org/10.1016/j.envsoft.2017.11.023

Abbreviations: ANN, Artificial neural networks; CBR, Case-based reasoning; EDSS, Environmental Decision Support Systems; FFNN, Feed-forward neural net; FCM, Fuzzy C-means; GK-FCM, Gustafson-Kessel fuzzy C-means; ICA, Independent component analysis; IWA, International Water Association; MLP, Multi-layer perceptron; ANFIS, Neural fuzzy inferences systems; ORP, Oxidation-reduction potential; PLS, Partial least squares; PCA, Principal component analysis; PCR, Principal component regression; QTA, Qualitative trend analysis; SOM, Self-organizing maps; SBR, Sequencing batch reactor; SVM, Support vector machines; SVR, Support Vector Regression; UPGMA, Unweighted Pair Group Method with Arithmetic Mean; WWTPs, Wastewater treatment plants; WNN, Wavelet neural nets.

industries before being discharged to a receiving water body. WWTPs are complex systems, which have to maintain high performance at all times, despite suffering from hourly, daily, and seasonal dynamics. WWTP operations have the particular feature that any "raw material", i.e. wastewater, must be accepted while the product, i.e. treated effluent, must adhere to its standards at all times. Furthermore, WWTPs have to adapt to new challenges posed by the society such as the removal of emerging pollutants, the minimization of greenhouse gases emissions, etc. (Hadjimichael et al., 2016). Overall, large amounts of data from WWTPs are being generated which need to be properly transformed into knowledge for enhancing their operation. Such knowledge can then be encapsulated into controllers or Environmental Decision Support Systems (EDSS) that allow maintaining high performance (and low emissions) at all times. During the last two decades, several driving forces that have intensified the development of computer-based techniques to transform data into knowledge in the wastewater treatment field.

The first driving force was control implementation, to increase the stability of the process ensuring good performance at all times, and to optimize the usage of resources (e.g. energy and chemicals). Control stimulated developments since the early 1970s (Olsson, 2012) on data cleaning, selection and transformation, which renders the data interpretable and useful for human inspection and automatic feedback control. Today, many sensors, such as those used to monitor dissolved oxygen, several nutrients, suspended solids and organic matter, have undergone important transformations, rendering them reliable and affordable (see Vanrolleghem and Lee, 2003; see manufacturers Hach[®], Endress + Hauser, S::CAN, etc.). The development of such sensors itself required the usage of data treatment methods (e.g. regression applied to information gathered from UV-VIS sensors). However, we realized that the installation of sensors and their maintenance efforts is insufficient to guarantee data quality and hence, methods were incorporated to allow for fast detection and diagnosis of faults. Also we incorporated methods to verify process normalcy and to create useful knowledge concerning plant malfunctioning and how to either improve plant performance or return it to normal operation. Hence, this stimulated the development of methods dealing with mass balances and data reconciliation for basic information extraction. Control development evolved from unit process control to sophisticated optimization and automation software packages, including rule-based systems and expert systems (Åmand et al., 2013; Ingildsen, 2002).

The second driving force was the transformation of data graveyards into data mines. It is evident that the incorporation of new challenges (from aeration control to system-wide control) and the increased levels of monitoring, control and supervision have led to the need for the handling of a large number of signals. Our current experience suggests that small WWTPs (~20,000 Population Equivalents, PE) can generate up to 500 signals (including online and offline signals), whereas larger ones (0.8–3 million PE) register analogical and digital signals exceeding 30,000 in number (Olsson et al., 2014; Freixó, 2016). As has been recognized, however, data-rich is all too often equivalent to information-poor (Nopens et al., 2007; Poynter, 2013). Indeed, vast amounts of data are languishing in databases, which are at best described as data graveyards and can certainly not be considered data mines. Indeed, current practice is arranged such that plant operators have an overwhelming stream of data at their hands, which is very difficult to process and analyse in a timely enough fashion to allow for better understanding or proper decision-making. As the effort to analyse data is costly because of a lack of trusted analytic data tools, potentially valuable information remains unavailable and unexploited (Yoo et al., 2008). Hence, methods appeared for Advanced *information extraction* to facilitate the interpretation of large datasets with multiple variables, i.e. multivariate methods such as principal component analysis (PCA), independent component analysis (ICA), and clustering. In addition, the large amounts of data stimulated the development of black-box models (such as artificial neural networks –ANN- or support vector machines –SVM-) which could be used for process optimization. Finally, other methods appeared for *Human-interpretable information extraction*, within the field of knowledge discovery (rule induction, decision trees, etc.) and management (ontologies). Taking advantage of increasing computing capacity, innovative knowledge-based systems have evolved to make use of both numerical models and heuristic knowledge in tandem with classical and innovative knowledgeacquisition techniques in EDSS.

The current data-rich, information-poor condition is a general problem that is not unique to the wastewater treatment industry. Indeed, many tools have been developed already and are popular within the chemical processing and paper and pulp sectors. Wastewater treatment operations are unique, however, for the following reasons. First, material inputs (i.e., wastewater) (1) cannot be stored in large quantities if the supply exceeds the process capacity (e.g., storm water), (2) cannot be discarded and ignored if they are of low quality (i.e., all discharged waters are accounted for in performance evaluations) and (3) are characterized by high temporal variability in both volume and quality. Therefore, borrowing methods from other engineering fields is not sufficient to guarantee the successful transformation of data into knowledge. The field of wastewater treatment requires specific adaptation of the methods to account for the uniqueness of the wastewater treatment process. The IWA (International Water Association) Instrumentation, control and automation conferences in particular have provided an excellent platform for such adaptations. A summary is found in Olsson (2012).

2. Literature review

2.1. Methodological approach

This section presents a review of peer-reviewed international journal papers that developed or evaluated techniques applied to WWTPs to maximize the potential of generated data and turn it into useful information and knowledge to improve wastewater treatment operation or support decision-making. The techniques that have been reviewed are organized according to the three following levels (Table 1): Basic information extraction, advanced information extraction, and Human-interpretable information extraction. Note that the levels of complexity (time required to compute a solution) increase gradually, ranging from the first (analysing single datasets or variables) to the third level (considering multiple variables and many types of knowledge). Within the advanced information extraction level, we have grouped the techniques as follows: the first group includes techniques based on dimension reduction; the second group includes techniques to infer linear functions from labeled training data; the third group describes techniques to infer non-linear functions; the fourth group consists of clustering techniques; and the last group includes Qualitative feature detection techniques. Within the humaninterpretable information extraction level we included rules, CBR, Fuzzy logic, EDSS and Ontologies. Hence, we mostly include data mining techniques, which come from an interdisciplinary field to discover patterns in large data sets and subsumes both machine learning, statistical modelling or visualization disciplines. A specific review of machine learning methods applied to the field of water and wastewater can be found in Hadjimichael et al. (2016). Mechanistic methods/models were excluded from this review.

Please cite this article in press as: Corominas, Ll.-g., et al., Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques, Environmental Modelling & Software (2017), https://doi.org/10.1016/j.envsoft.2017.11.023

Download English Version:

https://daneshyari.com/en/article/6961991

Download Persian Version:

https://daneshyari.com/article/6961991

Daneshyari.com