Environmental Modelling & Software 100 (2018) 291–301

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Detection and interpretation of anomalous water use for non-residential customers

Samitha Patabendige [a, b, *], Rachel Cardell-Oliver [a, b], Rui Wang [a, b], Wei Liu [a, b]

[a] CRC for Water Sensitive Cities, PO Box 8000, Clayton, Victoria 3800, Australia
[b] The University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia

A R T I C L E I N F O

Article history:
Received 12 May 2017
Received in revised form
18 September 2017
Accepted 18 November 2017
Available online 5 December 2017

Keywords:
Water demand management
Anomaly detection
Decision support
Smart meter

A B S T R A C T

Smart water meters can help businesses save water. But achieving this goal requires trusted algorithms for processing the data and intuitive, interactive software systems to support end users in decision making. This paper presents an algorithm and a web-based software system to detect and visualise anomalous water use. The algorithm calculates an anomaly score for each day, together with a rationale describing the symptoms of unusual water use. The score for each day is based on ten features of daily demand and its historical context. The score and its rationale are posted to users to help them track down the underlying physical causes of anomalies. Using data from two aquatic leisure centres, we demonstrate that anomaly scores give better coverage than traditional threshold-based systems, that end users are able to utilise the timely feedback to save water, and that the algorithm is reasonably robust to parameter settings.

Software availability

Name of Software: Smart Water Meter Anomaly Detection System
Developers: Samitha Patabendige, Rui Wang, Rachel Cardell-Oliver
Contact: samitha.maddumapatabendige@research.uwa.edu.au
Year first available: 2017
Hardware required: General-purpose Web server
Software required: Linux Server OS, Apache 2 Server, JRE 8, PHP 7, R 3.3.3
Languages: Java, Javascript, PHP, R
Availability: Contact authors
Case Study: http://datascience.csp.uwa.edu.au/AnomalySite
Support Languages: English
Licensing: GNU General Public License

1. Introduction

Reducing water consumption is a problem of growing importance because of the pressure on existing infrastructure caused by increasing urban populations, weather extremes and the rising costs of maintenance (Cominola et al., 2015). For example, one city water utility found that leaks occupy 28% of the total consumptions in commercial office buildings, 18% in shopping centres, and 22% in aquatic leisure centres (Best Practice Guidelines, 2007; Best practice guidelines, 2011). For businesses with high water use, such as aquatic leisure centres, a key goal is to identify and address unnecessary water use (Best practice guidelines, 2011). This paper addresses the problem of detecting unusual water use events, called anomalies, using smart water meters.

Timely, user-oriented alerts about anomalous water use help users to manage their demand more effectively. Common practices for monitoring water use are manual meter readings or using automated monitoring systems (Waterwise Council Program, 2017; Best Practice Guidelines, 2007; Best practice guidelines, 2011). Manually reading and recording meters is a laborious task and so systems that automate data collection and presentation are preferable.

Currently available automated systems can be classified into commercial or research systems. Commercial systems are designed to be general-purpose to suit different businesses. The majority of commercial systems provide basic statistics on consumption but do not detect more complex patterns of anomalous water use.

* Corresponding author.
E-mail addresses: samitha.maddumapatabendige@research.uwa.edu.au (S. Patabendige), rachel.cardell-oliver@uwa.edu.au (R. Cardell-Oliver), rui.wang@research.uwa.edu.au (R. Wang), wei.liu@uwa.edu.au (W. Liu).

https://doi.org/10.1016/j.envsoft.2017.11.028
1364-8152/

Additionally, systems such as Outpost[1] and Greensense View[2] have an alarm notification feature. For example, Outpost requires the user to choose thresholds and then an alert is sent automatically whenever consumption levels exceed the pre-defined thresholds. Although these systems eliminate the need for physical access to read water meters, they still require the involvement of experts who understand the business and are capable of setting thresholds and also re-tuning the thresholds for different contexts such as the change of seasons.

Monitoring systems built upon research outcomes are typically open-source and cost-free. Similar to commercial systems, most offer basic statistics on consumption. Research systems are scientifically oriented, so usually designed for specific purposes such as evaluating a particular algorithm. Nezhad et al. (Jarrah Nezhad et al., 2014) present *SmartD*: a smart dashboard that specifically aims to help data analysts visualise the data from smart meters. Liu et al. (2015) introduce *SMAS*: a smart meter data analytics system that focuses on analysing daily consumption profiles and how external factors, e.g. temperature, affect consumptions. Liu et al. (2016) also propose a software architecture *Lambda* designed for real-time anomaly detection on large data sets where the performance of the system is critical. *Lambda* requires special purpose hardware, i.e. clusters of servers. Janetzko et al. (2014) present an anomaly detection system with the focus on visualisation. The system provides different visualisations of multi-variate time series data for energy consumption but it does not report the rationale for its decisions.

Anomaly detection is a general approach in which machine learning algorithms are used to identify any unusual patterns in water meter readings, not just threshold alarms. Automatic anomaly detection is, however, a challenging task and existing methods suffer from a number of problems that hamper their application in the water domain. In the supervised machine learning approach an algorithm is trained using a large amount of historical data that has been labelled by an expert. However, such labelling is extremely time-consuming and the labelling has to be re-done for each new user (Nguyen et al., 2013). Unsupervised machine learning algorithms for anomaly detection are available (Rayana and Leman, 2016), but their results may have low accuracy in practice because anomalies are rare by definition, unexpected and also dependent on the context in which they occur, such as summer or winter, weekday or weekend. Another lack in existing anomaly detection methods is that they cannot explain transparently *why* demand is anomalous, and hence their results may not be trustworthy for business use. Finally, few anomaly detection methods are designed to integrate into existing management workflows, and so they are not applied in practice.

A distinctive feature of water used by businesses is its dependence on calendar contexts such as the season of the year, day of the week, opening hours and maintenance schedules. Anomaly detection systems need to be *calendar context-aware* of the relevant context for each day in order to avoid incorrect reports. Existing threshold-alarm systems need to be manually re-tuned: they are not context-aware. We propose a method for identifying business-relevant calendar contexts and for integrating these contexts into our anomaly detection algorithm.

Most previous smart metering studies focus on residential water users rather than the business users considered in this paper. But an important consideration for both types of users is how to integrate anomaly detection into the normal workflow of the user. We describe such an integrated system as a *workflow-aware* system. In a recent review Cominola et al. (2015) conclude that although there have been many studies on user-oriented residential demand management strategies, there has been limited integration between their specialized methodologies for the four phases of data gathering, characterising end-users, user modelling and personalised demand management systems. They conclude that there is "a clear need to shift research efforts from the development of specialized methodologies within each step of the procedure toward a more integrated approach that covers all the four phases" (Cominola et al., 2015). This paper addresses that gap.

The goal of the system presented in this paper is to develop and test a software system that provides automated, *calendar context-aware*, *workflow-aware* anomaly detection for water-using businesses such as leisure centres. The main contributions of this paper for achieving that goal are:

1. An automated machine learning algorithm that generates an anomaly score for each day of water use, with a rationale for that score. The algorithm is *calendar-context-aware*, which automatically take calendar contexts of businesses into consideration, eliminating the involvement from human experts for timely re-tuning thresholds.
2. A web-based software system that implements an active communication mechanism, which automatically reports detected anomalies to users via emails, and allows users to post feedback on the anomalies. The lightweight communication channel quickly establishes trustworthiness in users, enabling the system to be easily integrated into existing management work flows.

We have evaluated the system for two aquatic leisure centres with multiple swimming pools.

2. Anomaly detection algorithm

This section describes our method for calculating an anomaly score for each day of water use. Anomalies are data objects that deviate significantly from the rest of the data objects as they were generated by a different mechanism (Hawkins, 1980). In this study, the data are smart meter readings, which are typically presented as time sequence of consumption rates (e.g. average flow rate) during a fixed period for each day. For example, in a leisure centre, the water demand for a day is recorded as a vector of 96 values, measuring the average flow rate every 15 min. Intuitively, a smart water meter reading may be considered as anomalous when it is exceptionally high (or low) in volume compared to other readings. However, identifying a single reading is not of much interest to stakeholders, since it does not take into account its context, where the context is a set of data points that are expected to have similar volumes. So instead of analysing each single data point of a day, in this study, we aim to identify whether a day of water use is an anomalous based on the context of the time of day and relevant historical days of use.

Knowing *why* a given day has a high anomaly score helps non-residential water users to trace the underlying cause. Therefore, in addition to calculating an anomaly score of a day, our system gives a rationale for the score. For example, a day has an anomaly score of 0.75 because of an unusual flows between 6:00a.m. and 9:00a.m. Rationales summarise the symptoms of anomalous water use, which guides users in tracing physical causes.

Firstly, we identify the influential data points using calendar contexts. For example, water consumption is periodic, which changes by the seasons. Hence, using the past three month data of a given day as the context is more meaningful than using all historical data. Secondly, we identify features, such as the minimum and

[1] http://www.outpostcentral.com/.
[2] http://www.greensense.com.au/.