Environmental Modelling & Software 100 (2018) 302–313

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A non-intrusive approach for classifying residential water events using coincident electricity data

J.S. Vitter [a, *], M.E. Webber [a, b]

[a] Department of Mechanical Engineering, The University of Texas at Austin, 204 East Dean Keeton Street Stop 2200, Austin, TX 78712, United States
[b] Energy Institute, The University of Texas at Austin, 2304 Whitis Ave Stop C2400, Austin, TX 78712, United States

ARTICLE INFO

Article history:
Received 11 May 2017
Received in revised form
21 August 2017
Accepted 18 November 2017

Keywords:
Water end use event
Water disaggregation tool
Residential water and electricity data
Support vector machine
Confusion matrix

ABSTRACT

This study evaluated the potential for circuit-level electricity data to improve performance by a water end-use disaggregation tool. Support vector machine classifiers were employed to categorize observed water events from an extensive dataset published in the literature. Additional electricity-related event features were assigned depending on temporal proximity to recent clothes washer or dishwasher events. Classifiers were trained on a portion of the dataset with and without the electricity-related features, then tested on an equally sized portion of the dataset. A classifier also categorized events from the testing dataset where event durations were adjusted to match larger sampling intervals, from 10s up to 120s. Specific electricity-related features significantly improved classifier performance for clothes washer, dishwasher, and shower events. Classifier performance was maintained for longer events as sampling frequency decreased, although performance for short duration events decreased. Overall, these results indicate significant potential benefits from integrating electricity-related features for water disaggregation tools.

1. Introduction

As cities embrace smarter management systems for energy and water resources, advanced metering infrastructure (AMI) has become a common enabler for water utilities to improve their business models, prevent water theft, detect leaks, and improve customer service (Hawkins and Berthold, 2015). Due to energy requirements for water processes, increased monitoring for end uses of water to prevent losses or encourage conservation can also have non-trivial energy impacts (Carlson and Walburger, 2007; Sanders and Webber, 2012). These factors, along with projected growth for water demand in urban areas (Davies, 2016; Bruun et al., 2017) motivate efforts to better characterize and understand the time-varying nature of urban water demand. With the development of sophisticated water meters and high resolution sensors, water use from entire homes can be non-intrusively metered, with water consumption sampled on a sub-daily basis (Cominola et al., 2015). With certain meters or data loggers, water use can be measured with very high temporal resolution corresponding to sampling intervals of a few seconds (DeOreo et al., 1996). High resolution water data has enabled applications that characterize patterns of appliance or fixture water consumption, and use these patterns to categorize whole-home water data by end uses (Cominola et al., 2015). An early example of end-use disaggregation software, *TraceWizard*, employed a decision tree algorithm that uses household audits of appliance stock, pre-defined templates, and trained analysts to predict water events (Mayer and DeOreo, 1999; DeOreo, 2011; DeOreo et al., 2016). Other past examples of disaggregation tools include *Identiflow*, *HydroSense*, and *Autoflow* (Cominola et al., 2015). *Identiflow* used decision trees to disaggregate water consumption semi-autonomously at the household level (Kowalski and Marshallsay, 2003). *HydroSense* used a pressure sensor to identify unique pressure waves, then implemented a Bayesian approach to identify when individual fixtures were opened and closed (Froehlich et al., 2009, 2011). A more recent piece of software, *Autoflow*, was developed using an echelon of methods and machine learning algorithms such as hidden Markov models, dynamic time warping, gradient vector filtering, and artificial neural networks (Nguyen et al., 2013a, b, 2014, 2015). *Autoflow* was developed and validated using recent end-use studies conducted in Brisbane and Melbourne, Australia (Beal and Stewart, 2011, 2014).

* Corresponding author.
E-mail addresses: scott.vitter@utexas.edu (J.S. Vitter), webber@mail.utexas.edu (M.E. Webber).

https://doi.org/10.1016/j.envsoft.2017.11.029
1364-8152/

The *TraceWizard* software has been used widely in the past for large end-use studies, and *Autofow* is being developed for autonomous commercial applications. Although the methods used by these two approaches are significantly different, both operate using water consumption data with very high temporal resolution (Cominola et al., 2015). For example, the *Autoflow* algorithm was developed with and validated on five-second interval data with a precision of 0.014 L per pulse from the data logger (Nguyen et al., 2013a, b, 2015), and the TraceWizard algorithm uses 10-s interval data with approximately 0.01 gallons per pulse from the data logger. (DeOreo et al., 1996; Mayer and DeOreo, 1999; DeOreo et al., 2016). While high resolution data enables accurate disaggregation, the relationship between sampling frequency and classification accuracy has not been widely reported by the existing literature. This relationship is related to the Nyquist sampling theorem, which states that the highest frequency event that can be measured is at best one-half of the sampling frequency (Matthews et al., 2008). To capture "microscopic" features of a signal, rules of thumb suggest that up to 20 data points might be required per fundamental period (Zeifman and Roth, 2011).

A similar blind identification problem has been studied in the electricity sector, where Non-Intrusive Load Monitoring (NILM) algorithms attempt to identify distinct appliance signatures from aggregate energy data (Kolter et al., 2010, Kolter et al., 2012; Kolter and Ferreira, 2011; Zeifman and Roth, 2011; Zoha et al., 2012; Cominola et al., 2017). In general, pattern recognition based approaches are preferred over optimization approaches (Zoha et al., 2012). Such algorithms might be relevant to the water event disaggregation problem, although their portability has not been assessed (Cominola et al., 2015).

When both electricity and water consumption data are collected, a combined approach could use established NILM algorithms to identify electrical appliance signatures of water using devices, and proceed to use the information about electricity consumption to improve water disaggregation methods. Unlike appliance-level water sensors, appliance-level electricity data of major appliances (i.e. clothes washer, dishwasher, electric water heater) can be easily metered via individual circuits. Circuit-level electricity data can simplify a combined approach, allowing water disaggregation methods to use measured electricity consumption by major appliances rather than predictions from NILM algorithms. Electricity consumption data has not been incorporated into existing water disaggregation tools, potentially due to the unique datasets required for analysis. As more homes adopt smart meters for both electricity and water, however, there is significant potential for synergetic applications of both types of data. A recent analysis employed clustering techniques on circuit-level clothes washer and dishwasher data collected from homes in Central Texas to identify distinct cycle patterns (McCartney, 2016). Once the cycles were clustered, water consumption for a given cycle was estimated as the median of water use during cycles from that cluster. Excess water consumption during each cycle was assigned to other end-uses using an approximation (McCartney, 2016). This analysis was preliminary, and water consumption could not be validated against a known dataset.

To date, the authors are unaware of a water disaggregation tool that combines data for electricity consumption and water consumption to make predictions against a known dataset. This study is a first step in that direction. The purpose of this analysis is to evaluate if circuit-level electricity data from clothes washer and dishwasher appliances can improve water event classification accuracy, and if performance is maintained as the temporal resolution of water data decreases (sampling interval increases). To minimize intrusiveness, data from the Residential End Uses of Water Study (REUWS) Version 2 (DeOreo et al., 2016) were used to provide ground truth information to train classifiers. Features related to appliance electricity consumption were defined for each event in the REUWS dataset according to their temporal proximity to known clothes washer or dishwasher events. Support vector machine (SVM) classifiers were then trained on a portion of the original dataset with and without electricity-related features to assess if circuit-level electricity data can significantly improve classifier performance. Classifier performance was also evaluated for sampling intervals as large as two minutes to explore the trade-off between classifier performance and the temporal resolution of data sampling. A flowchart to visualize the components of the study is shown in Fig. 1.

This study is novel because it represents the first attempt to incorporate information about simultaneous electricity consumption into a water end-use disaggregation tool. The results from this study provide an indication of the potential for electricity data to support classifier performance even when the temporal resolution of water data is diminished. In the future, water and electricity disaggregation tools might be replaced by low-cost direct sensing of individual fixtures or appliances in a manner that integrates seamlessly into a smart-home environment. However, it is unclear if or when direct sensing will become a standard component of new buildings. In the interim, water disaggregation tools can provide useful information about consumption at the fixture-level using household-level data from smart meters, whose adoption is ongoing. As smart-home concepts receive more attention, this

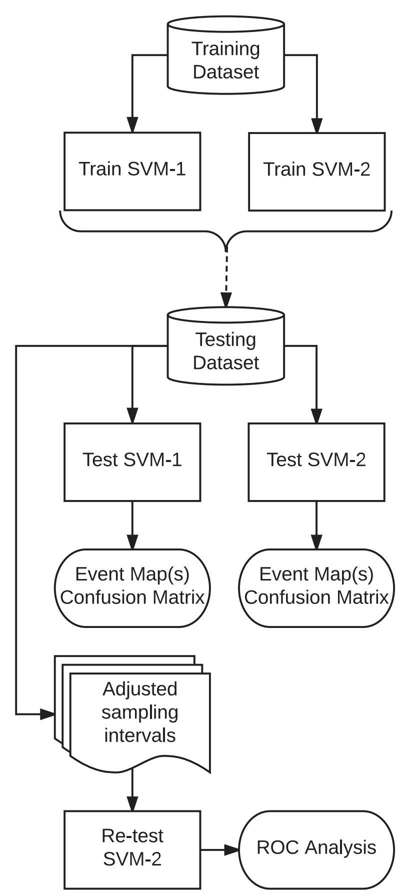

Fig. 1. After datasets are assembled, support vector machines (SVM) are used to train classifiers and test their performance. Outputs of the analysis include event maps, confusion matrices to indicate classifier performance, and receiver operator characteristics (ROC) to assess the tradeoff between classifier performance and data sampling interval.