Empowering Personalized Feedback on Hot Water Usage: A Field Study with Shower Meters

Sebastian A. Günther
Information Systems and Energy Efficient Systems, University of Bamberg
Bamberg, Germany
sebastian.guenther@uni-bamberg.de

Carlo Stingl
Information Systems and Energy Efficient Systems, University of Bamberg
Bamberg, Germany
carlo.stingl@uni-bamberg.de

Vlad C. Coroamă
Department of Computer Science, Institute for Pervasive Computing, ETH Zurich
Zurich, Switzerland
vlad.coroama@inf.ethz.ch

Samuel Schöb
Mechanical Engineering, Massachusetts Institute of Technology
Cambridge, USA
sschoeb@mit.edu

Thorsten Staake
Information Systems and Energy Efficient Systems, University of Bamberg
Bamberg, Germany
thorsten.staake@uni-bamberg.de

ABSTRACT

Digitalization enables an ever-increasing opportunity to promote resource conservation by providing timely consumption feedback to individuals. Yet, in multi-person households, appliances and fixtures are often shared, which makes it difficult to deliver person-specific feedback for each user. In this paper, we tackle this problem for household's most energy-intensive day-to-day activity – showering – by leveraging granular water consumption data. To this end, we collected labeled time series data of 691 shower events from 28 individuals, defined features, implemented and adjusted several classifiers, and analyzed the feasibility of our identification approach. Across all locations, the results – which were evaluated with stratified five-fold cross validations – provide robust evidence that the presented approach can indeed identify users reliably immediately after the end of a shower event. More specifically, the classifier with the best overall performance (Random Forest) achieved an average accuracy of 83.2% even for the most challenging environment in our field test (differentiating between five individuals in a company shower) and reached an average accuracy of 98.8% for a two-person household. Moreover, the approach requires only little training data for a satisfactory performance.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; • **Information systems →** *Information systems applications*;

KEYWORDS

User Identification, Personalized Feedback, Hot Water Consumption, Machine Learning

ACM Reference Format:
Sebastian A. Günther, Carlo Stingl, Vlad C. Coroamă, Samuel Schöb, and Thorsten Staake. 2020. Empowering Personalized Feedback on Hot Water Usage: A Field Study with Shower Meters. In *The 35th ACM/SIGAPP Symposium on Applied Computing (SAC '20), March 30-April 3, 2020, Brno, Czech Republic.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3341105.3374053

1 INTRODUCTION

With the ubiquity of digital technologies, more and more behavior data about everyday activities become available. Driven by the recent political aims to reduce the impact of climate change by demand-side management, many approaches leveraged such data to provide individuals with an understandable form of resource use to promote behavioral change. In this context, feedback was shown to be most effective in fostering resource conservation when it is specific, delivered in real-time, and at the place of action (e.g., [7, 19]), which stresses the importance of digital technologies. Recent research has shown that feedback can be effective even without financial incentives and in the absence of selection effects (i.e., [20]).

In fact, in recent years a vivid research domain has developed to make feedback more specific. For instance, many approaches have been proposed that leverage electricity data from a single sensor (e.g., a smart meter) to recognize the energy consumption of individual appliances. Strikingly, Google Scholar counts 2,580 papers for the related term "energy disaggregation" of which 69% were published in the past four years. In the same vein, several approaches were proposed that allow monitoring of the water consumption at the appliance and fixture level by using different types of sensors and algorithms (e.g., [3, 11, 16]). Little research, however, has tried to provide individuals at home with person-specific feedback on their use of shared appliances or fixtures. Existing work is mostly based on tracking of individuals within their home (e.g., [5, 15]), which is most likely not feasible for a large-scale roll-out. Similarly, periodically prompting individuals to log their everyday usage of

ACM ISBN 978-1-4503-6866-7/20/03.
https://doi.org/10.1145/3341105.3374053

shared appliances and fixtures is unlikely to work. Interestingly, research in other domains has shown that individuals can be identified by behavioral data alone (e.g., [14, 21]), which indicates that feedback systems can potentially automatically identify users for person-specific feedback.

We add to the discussion an energy-intensive day-to-day activity where no personalization work has been published that relies solely on consumption data: Showering. We target showering for user identification for three reasons. First, a typical shower consumes around 45 liters of water and 2.6 kWh of heat energy [19], with a high carbon intensity due to the widespread usage of fossil sources for hot water generation [8]. For comparison, the average EU household uses 1 kWh a day for lighting [13]. Second, shower installations in households are often shared by multiple persons, thus resembling a natural setting to disaggregate water consumption and to cost-effectively induce resource conservation. Lastly, resource consumption in the shower varied substantially between individuals of a previous study [19], which indicates the necessity of user differentiation for personalized feedback strategies.

Although activity-specific real-time feedback on showering fostered savings of 22% in a household setting (i.e., 0.46 kWh savings per shower), the feedback has been given in a general manner and it was not aligned with the individuals' past consumption history, personality and preferences [19]. By contrast, literature suggests that feedback needs to be designed towards the individual to unleash its full potential ([6, 17, 18]). Similarly, due to the lack of user identification, the feedback was bound to the place of action, thus, not allowing for subsequent motivating elements (i.e., peer comparisons on a smartphone) or user-specific content (i.e., user-specific conservation goals) in multi-person households. We believe that such personalized feedback systems can enhance the existing conservation effects, however, it requires the ability to automatically identify individuals, since asking users after each consumption event for their identity is not convenient and not reliable: Although we have explicitly requested this for the paper's experiment, the individuals in the private setting only indicated their identity for 35% of their showers.

2 IDENTIFICATION APPROACH

In order to evaluate the feasibility of user identification, we collected ground truth data in five private households near Nuremberg (Germany) with varying number of residents. To explore the limitations of our approach, we additionally collected data in one company in Zurich (Switzerland) where, according to the staff, more than 20 employees regularly use the company's showers.

For the data collection, we deployed shower meters which measure water and energy consumption. Whenever water flows, the current resource consumption state is communicated via Bluetooth with a data granularity of approximately two samples per second. Each data point contained aggregated measures (e.g., water used until that measurement point) as well as current properties of the shower (i.e., current temperature or flow rate). Due to this high data resolution and associated storage limitation of the shower meters, we additionally relied on Raspberry Pis collecting the measurements live via Bluetooth from the shower meter and transmitting them to our server infrastructure.

Besides collecting consumption data, we relied on the individuals logging their usage of the shower. While the residents of the households took notes in form of logbooks, the employees of the company used a mobile app for this purpose. Both forms of logging required the individuals to provide us with the combination of a personal identifier and a timestamp. This information was needed to match the respective recorded shower data with the personal identifier for the creation of the ground truth data set. In the private setting, only one shower installation was present and each logbook entry had to be mapped to only one data source (shower meter), whereas in the corporate setting several shower installations were possibly used at the same time. Therefore, the participants in the corporate setting had to indicate the shower installation they used as well.

In order to identify individuals based on their shower usage, it is necessary to pre-process the data (i.e., match the log entries to the respective shower event), and to derive meaningful features for high predictive power, which we outline in the following. We derived 34 features from the granular data, which can be roughly divided into three categories. The first category represents temporal aspects of the shower (day of the week or time of the day) which might help to distinguish individuals based on their daily routines even when their showering behavior is similar. The second category represents statistical measures (e.g., mean temperature, mean flow rate, duration of the shower) which should relate to the shower preferences of the individual. The third category contains more advanced features that were extracted from the granular data (e.g., number of water flow stops per shower, a measure for the number of temperature adjustments per shower).

3 EVALUATION

In this section, we present the results of our identification approach and briefly discuss the relevance of training data to achieve a satisfactory performance.

3.1 Evaluation setting

We approach the identification problem by using a set of standard algorithms. For the supervised classification methods we consider the white-box-methods Decision Tree (DT) and Lasso Logistic Regression (LogReg) [9, 12], as well as the black-box methods Artificial Neural Network (NN), Random Forest (RF) and Support Vector Machine (SVM) [2, 4, 22] using the R package "mlr" [1]. For each of the shower meters, we define a separate classification problem for two important reasons. First, the characteristics such as the flow rate or the water heating system vary substantially among different shower installations. Using one classification problem for all the labeled showers would most likely bias the results: Instead of learning consumption characteristics of individuals, it would probably just recognize the characteristics of the shower installation. Second, the identification performance in a real-world residential scenario would typically imply differentiating between a relatively low number of individuals and not more than 25 individuals as in the present study. Going for a single classification problem, there is a higher chance that multiple individuals share the same consumption habits (i.e., time of showering, temperature preference, number of water flow stops), leading to an underestimated performance of the classifiers for a real-world scenario. Moreover, even though the individuals in the corporate setting could have used different shower installations from time to time, they used almost always the same installation. As a consequence, the data set is also too restricted to investigate whether we can identify users based on the data from a different shower installation.

For all main results we have chosen a five-fold stratified cross validation which splits the available data randomly in five mutually exclusive sets. Furthermore, we have repeated the cross validation ten times enabling us to analyze the stability of the identification. We base the identification reliability on two performance metrics: Accuracy (i.e., the proportion of correct predictions) and area under the curve (AUC) adopted to multiclass problems as follows:

$$AU1P = \frac{1}{c\,(c-1)} \sum_{j=1}^{c} \sum_{k \neq j}^{c} p(j)\, AUC(j,k) \tag{1}$$

where c is the number of classes, $p(j)$ is the a priori distribution of class j and $AUC(j,\ k)$ is the AUC between the classes j and k [10]. In the following, when we describe the classifier performances in terms of AUC, we refer to this multiclass definition.

Due to our five-fold stratified cross validation, we need at least five observations per individual to estimate the AUC. As a consequence, we limit the data set to individuals with at least five showers (i.e., 28 individuals and 691 showers). To not introduce any selection biases resulting from the evaluation of this subset of individuals, we made additional robustness checks. Overall, the analysis is robust in terms of accuracy and AUC when including again the individuals with less than five log entries (when evaluating a cross validation with less than five-folds).

3.2 Evaluation results

Figure 1 shows the mean accuracy and the mean AUC values of our five classifiers across the different households and the corporate setting. For the accuracy, we set as baseline the *Majority class* classifier which maximizes performance by predicting the meter-specific user with the most labelled showers. Applied on the data set, the *Majority class* classifier achieves values between 24.1% (Household 4) and 58.3% (Household 1).

For all of the shower installations, the regular classifiers perform better than the *Majority class* classifier, thus indicating that they truly learned user-identifying patterns from the consumption data. Strikingly, in the first household with two individuals, all the classifiers reached an accuracy of at least 93.5% (NN), with a maximum of 98.8% (RF). Next, we analyze the results of household 4 with labelled showers from five individuals. Similar to household 1, the classifiers reach remarkable accuracy values with a maximum of 95.1% (RF). Only the DT falls significantly in performance, but still achieves satisfactory accuracy. By contrast, the performance in household 5 is slightly worse (minimum accuracy 56.2%, maximum accuracy 83.3%). With only 30.7% shower events as compared to household 4, this potentially indicates the need for more training data to achieve a similar performance level. Notably, our identification approach works in the corporate setting similar as good as in the private setting when comparing the performances of the regular classifiers with those of the *Majority class* classifier. Finally, the analysis of the AUC values indicates that the RF and the LogReg perform on average the best.

Next, we analyze the relevance of training data to achieve satisfactory results for user identification. In a real-world scenario,

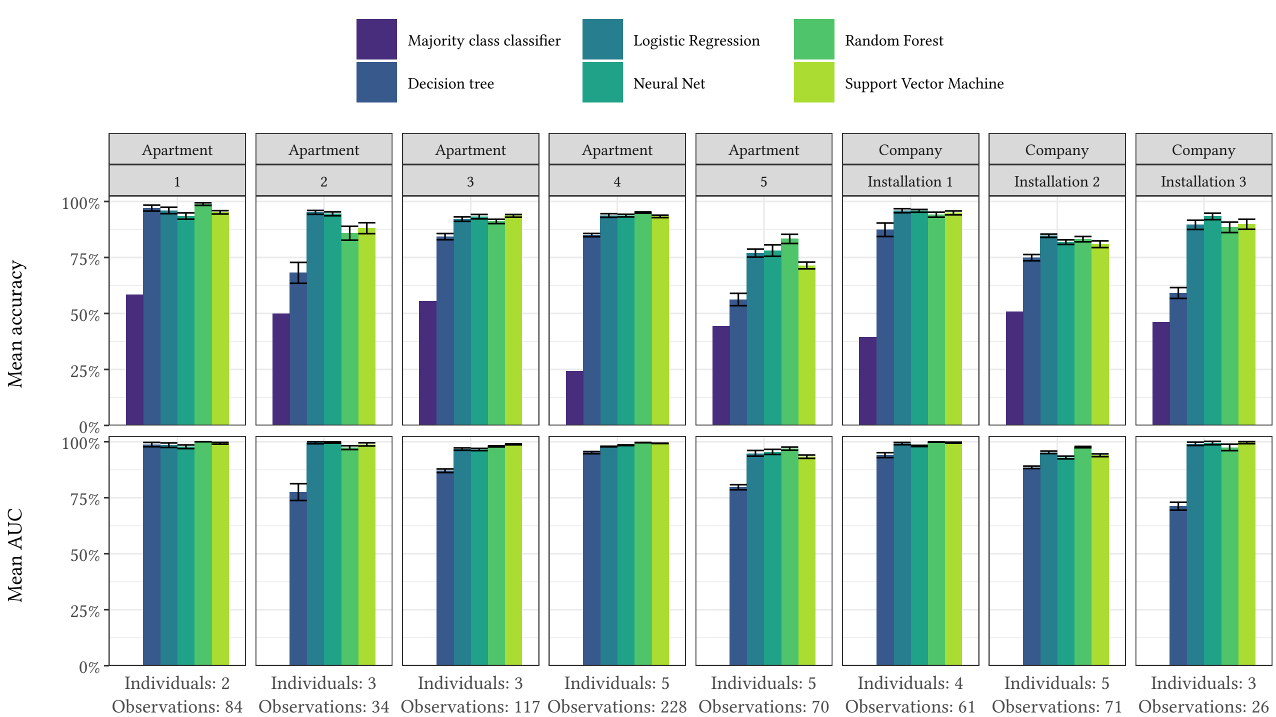

Figure 1: Performances of the classifiers. Error bars indicate the 95% confidence interval of the five-fold cross validations

individuals are most likely not willing to label a lot of their showers to get the system running. A requirement of the system is, thus, to work with a few training instances per user.

To this end, we benchmark the classifiers by limiting artificially their training set and determining subsequently their classification performances. Surprisingly, this additional analysis indicates that many classifiers perform well with little training data available. More specifically, with only one training instances per individual, the RF was already better than the *Majority class* classifier across all shower installations. Still, we recommend collecting three to four training instances per individual before applying the approach in practice. Consequently, we provide evidence that such a system can be deployed even when the individuals are not willing to log many of their showers to setup the system.

4 CONCLUSION

In this paper we have presented and evaluated an approach that leverages granular water consumption data in order to identify users after an energy-intensive everyday activity (showering). Considering the performances across different settings, our approach achieved very satisfying results: We have demonstrated that individuals can be reliably detected after the end of a shower. For instance, the RF classifier achieved an overall average accuracy of 90.8% in the households, differentiating on average between 3.6 individuals. In the corporate setting, it reached an overall average accuracy of 88.6%, differentiating on average between 4 individuals. The RF classifier's accuracy values were constantly much better than the *Majority class* classifier. In addition, only very little training data is needed to setup such a user identification system.

We have successfully developed an approach that makes it possible to identify consumers of hot water usage in the shower. The results are promising and encourage future research in the domain of feedback interventions. Due to advances in energy efficiency, human behavior is an increasingly important factor in the end energy use. As a consequence, smart sensors are an important lever to induce resource conservation by giving feedback on resource consumption. Identifying users and providing personalized feedback can serve as an additional way to enhance the individual curtailment efforts in curbing CO_2 emissions. Furthermore, it can enable future services to increase comfort and help researchers to better understand the behavior of individuals.

ACKNOWLEDGMENTS

This work has been funded in part by the Bavarian State Ministry of Education, Science and the Arts within the framework "TechnologieAllianzOberfranken (TAO)". The authors gratefully thank Regina Schmitt for her help in carrying out the study, the Ergon AG for making the research in the corporate setting possible and especially the participants for their invaluable support.

REFERENCES

[1] Bernd Bischl, Michel Lang, Lars Kotthoff, Julia Schiffner, Jakob Richter, Erich Studerus, Giuseppe Casalicchio, and Zachary M Jones. 2016. Mlr: Machine Learning in R. *Journal of Machine Learning Research* 17, 170 (2016), 1–5.

[2] Leo Breiman. 2001. Random Forests. *Machine learning* 45, 1 (2001), 5–32. https://doi.org/10.1023/A:1010933404324

[3] Davide Carboni, Alex Gluhak, Julie McCann, and Thomas Beach. 2016. Contextualising Water Use in Residential Settings: A Survey of Non-Intrusive Techniques and Approaches. *Sensors* 16, 5 (2016), 738. https://doi.org/10.3390/s16050738

[4] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology* 2, 3 (2011), 27:1–27:27.

[5] Yun Cheng, Kaifei Chen, Ben Zhang, Chieh-Jan Mike Liang, Xiaofan Jiang, and Feng Zhao. 2012. Accurate Real-Time Occupant Energy-Footprinting in Commercial Buildings. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM Press, Toronto, Ontario, Canada, 115. https://doi.org/10.1145/2422531.2422552

[6] Dora L Costa and Matthew E Kahn. 2013. Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. *Journal of the European Economic Association* 11, 3 (2013), 680–702. https://doi.org/10.1111/jeea.12011

[7] Karen Ehrhardt-Martinez, Kat A Donnelly, and Lattner, S. 2010. *Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities*. Technical Report ACEEE Research Report E105. American Council for an Energy-Efficient Economy. 140 pages.

[8] European Statistical Office. 2019. Energy Consumption in Households. https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households#Energy_products_used_in_the_residential_sector.

[9] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research* 9 (2008), 1871–1874.

[10] C. Ferri, J. Hernández-Orallo, and R. Modroiu. 2009. An Experimental Comparison of Performance Measures for Classification. *Pattern Recognition Letters* 30, 1 (2009), 27–38. https://doi.org/10.1016/j.patrec.2008.08.010

[11] Sara Fontdecaba, José A. Sánchez-Espigares, Lluís Marco-Almagro, Xavier Tort-Martorell, Francesc Cabrespina, and Jordi Zubelzu. 2013. An Approach to Disaggregating Total Household Water Consumption into Major End-Uses. *Water Resources Management* 27, 7 (2013), 2155–2177. https://doi.org/10.1007/s11269-013-0281-8

[12] Torsten Hothorn, Kurt Hornik, and Achim Zeileis. 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15, 3 (2006), 651–674. https://doi.org/10.1198/106186006X133933

[13] Lapillonne, B., Pollier, K., and Samci, N. 2015. Energy Efficiency Trends for Households in the EU. https://www.odyssee-mure.eu/publications/efficiency-by-sector/households/household-eu.pdf.

[14] Robert J. Orr and Gregory D. Abowd. 2000. The Smart Floor: A Mechanism for Natural User Identification and Tracking. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*. ACM, The Hague, The Netherlands, 275–276. https://doi.org/10.1145/633292.633453

[15] Juhi Ranjan, Erin Griffiths, and Kamin Whitehouse. 2014. Discerning Electrical and Water Usage by Individuals in Homes. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM Press, Memphis, Tennessee, 20–29. https://doi.org/10.1145/2674061.2674066

[16] Samuel Schöb, Sebastian A. Günther, Karl Regensburger, and Thorsten Staake. 2018. NIWM: Non-Intrusive Water Monitoring to Uncover Heat Energy Use in Households. *Computer Science - Research and Development* 33, 1 (2018), 127–133. https://doi.org/10.1007/s00450-017-0353-8

[17] P. Wesley Schultz, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science* 18, 5 (May 2007), 429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x

[18] Bernadette Sütterlin, Thomas A. Brunner, and Michael Siegrist. 2011. Who Puts the Most Energy into Energy Conservation? A Segmentation of Energy Consumers Based on Energy-Related Behavioral Characteristics. *Energy Policy* 39, 12 (2011), 8137–8152. https://doi.org/10.1016/j.enpol.2011.10.008

[19] Verena Tiefenbeck, Lorenz Goette, Kathrin Degen, Vojkan Tasic, Elgar Fleisch, Rafael Lalive, and Thorsten Staake. 2018. Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation. *Management Science* 64, 3 (2018), 1458–1476. https://doi.org/10.1287/mnsc.2016.2646

[20] Verena Tiefenbeck, Anselma Wörner, Samuel Schöb, Elgar Fleisch, and Thorsten Staake. 2019. Real-Time Feedback Promotes Energy Conservation in the Absence of Volunteer Selection Bias and Monetary Incentives. *Nature Energy* 4, 1 (2019), 35–41. https://doi.org/10.1038/s41560-018-0282-1

[21] Wei Wang, Alex X. Liu, and Muhammad Shahzad. 2016. Gait Recognition Using Wifi Signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM Press, Heidelberg, Germany, 363–373. https://doi.org/10.1145/2971648.2971670

[22] Andreas Zell, Niels Mache, Ralf Hübner, Günter Mamier, Michael Vogt, Michael Schmalzl, and Kai-Uwe Herrmann. 1994. SNNS (Stuttgart Neural Network Simulator). In *Neural Network Simulation Environments*, Josef Skrzypek (Ed.). Vol. 254. Springer US, Boston, MA, 165–186. https://doi.org/10.1007/978-1-4615-2736-7_9