Non-intrusive disaggregation of water consumption data in a residential household

Ilya Kozlovskiy[1], Samuel Schöb[2], and Mariya Sodenkamp[3]

Abstract: The water conservation campaigns in residential households are hindered by the poor understanding of residents of how much water they use. For the better designed interventions new tools are necessary to educate the consumers on the water usage of different consumption events. In this paper we use the fine grained (0.5 Hz) water consumption data that was collected non-intrusively in a household over the period of 21 days to develop such tools. We examine the collected data and disaggregate the consumption events into three different categories: short events (e.g., toilet flush), long regular events (e.g., washing machine) and long irregular events (e.g., showers). To achieve this, we use clustering methods, based on level set trees, to identify groups of events that are similar to each other.

Keywords: Water conservation, Disaggregation, Water consumption, Smart meter

1 Introduction

An escalating demand on potable water resources resulting from increasing populations, droughts and unpredictable weather patterns due to climate change is commonplace in many parts of the world [Ba08]. As a result, the sustainable management of urban water has become imperative, particularly for countries prone to severe droughts [Wi11]. In the USA domestic (residential) water use was the third largest water use category after thermoelectric power generation and irrigation [Bu16]. The residential sector is the largest urban water use sector, and it offers the largest volume of potential savings compared with other urban sectors [Gl03]. But most individuals have poor understanding of how much water they use [At14]. They only get the annual, or in best case monthly bills, and lack the tools to monitor and improve their water consumption [Ta15]. The problem of residents not getting the immediate feedback also occurs in the energy sector with the electricity consumption. In this case multiple non-intrusive load disaggregation methods were developed to better educate the residents on their consumption [Ha92, Zo12, Do13].

A study that tried to educate the residents on their water expenditure with the help of in home displays has only reached a short term benefit [Fi13]. On the other hand, studies displaying the water consumption of individual showers have produced substantial (~22%) reduction in short [Ti13] and long term [Ta15] water consumption. But this

[1] Universität Bamberg, Kapuzinerstraße 16, 96047 Bamberg, ilya.kozlovskiy@uni-bamberg.de
[2] Universität Bamberg, Kapuzinerstraße 16, 96047 Bamberg, samuel.schoeb@uni-bamberg.de
[3] Universität Bamberg, Kapuzinerstraße 16, 96047 Bamberg, mariya.sodenkamp@uni-bamberg.de

method required the installation of a metering device directly into the shower. To create intervention campaigns more suitable for the mass market we therefore look into possibility of ***non-intrusive disaggregation of water consumption.*** Our aim is to show that it is possible based on the overall water consumption of the households to provide the residents with the information on how much water they consume in individual categories (e.g., showers, toilets, washing machines, or faucets).

2 Data collection methodology

To collect the water consumption data, we have mounted two sensors directly to the water intake pipe of multiple households. In this study we only consider the data from one household. The first sensor measures the flow speed in the pipe in m/s and with the measured interior diameter of the pipe we could calculate the volumetric flow rate also in l/min. The second sensor measured the temperature. Both measurements were collected simultaneously on average once every two seconds. The measurements were collected over the period of 21 days from 17.04 to 08.05.2016. Both of the sensors produce the measurements with a small error. The values for the flowrate are only influenced, by the water consumption in the household, but the values for the temperature also change during the day.

3 Three Step Disaggregation of Water Consumption Data

In order to identify the different classes of consumption patterns we have developed a three step methodology for analysis of water consumption. As a first step we identify the times where the active water withdrawal takes place (water consumption events) by determining the times with no consumption. In the second step we identify and describe the events that occur often (e.g., toilet flushes have similar consumption pattern). And in the last step, we identify and describe the events that are more rare, but use more water (i.e., events with longer duration like showers).

The sensor data for the water flow is the primary evidence to determine if the water is consumed. Since the sensor data is noisy, it is important to detect the level during which water withdrawal is highly probable. To do this, we compute the kernel density estimation over all sensor values for velocity and take the first local minimum as the boundary. We use this method since we assume, that the error in sensor measurements is normally distributed. The kernel density estimation is represented in the Fig. 1. Here we set the boundary value to 0.13 l/min. Every value for velocity that lies below this boundary is then set to zero. With this, we can define individual water extraction events, as time segments with the positive flow velocity. In the considered household we detect 1807 individual water extraction events during the 21-day period. Most of the events have only a short duration (median duration is 18 s) and nearly constant (up to measurement error) water flowrate. During most of the events there is also a slight drop (average 0.05°) in the temperature. Here we also make the assumption that all the events are singular and not a superposition of multiple withdrawals. Since most of the withdrawal events are short, this assumption should be satisfied in nearly cases.

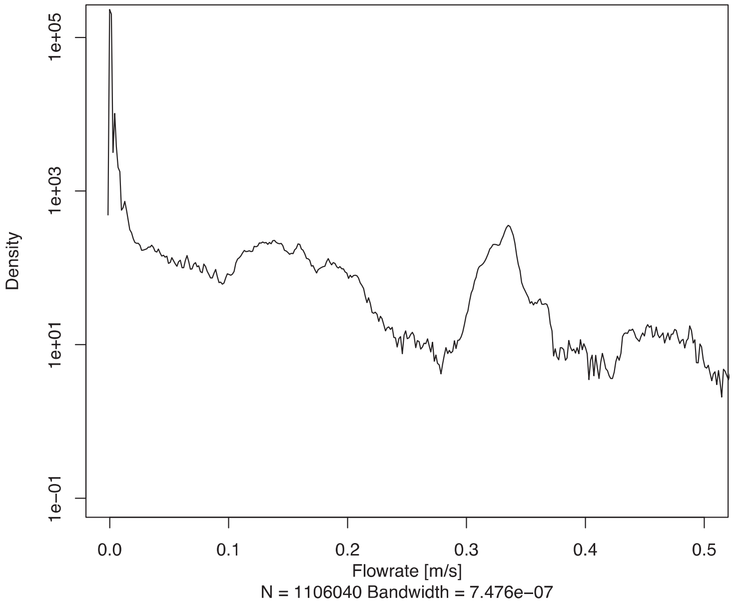

Fig. 1: The kernel density estimation of the water velocity sensor values in one household in 21-day period

Since most of the events have a similar pattern to each other we need to define measures that describe the events more closely. The definition of these features is an engineering task, in this work we use the following 5 features: duration of the event, average flowrate during the event, time of the day (in seconds), time with water consumption during the 20-minute window surrounding the event, and the temperature change during the event. The temperature change is calculated as the difference between the mean temperature during ten measurements before and after the event. Our goal is to use the computed features to find groups of similar events that occur often (i.e., the the density is higher for these events). Therefore, we use the density-based clustering algorithm that relies upon the level set trees [Ha75, St11, Ke13]. In this algorithm the individual clusters are the simple connected components at the given density levels. The resulting clusters are then presented in a tree hierarchy allowing for easy interpretation and visualization of the results. In our approach, we use the implementation in the R-package TDA [Br15], with the Gaussian kernel density estimator, the bandwidth h of 0.1 and k=100 for neighborhood estimations. The resulting hierarchy tree is presented in the Fig. 2. The individual clusters are the colored vertical lines and their length represents for how long the clusters survives until it connects into other clusters. We find 2 main clusters in our dataset, each with two sub-clusters. For better visualization we plot the event duration vs. the average flowrate with the color corresponding to the cluster in the Fig. 3.

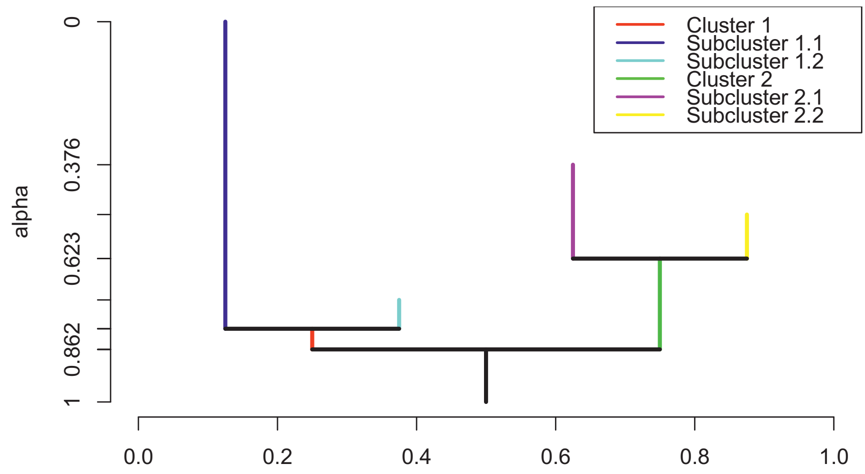

Fig. 2: The sublevel tree with different clusters (vertical lines) as result of the density clustering

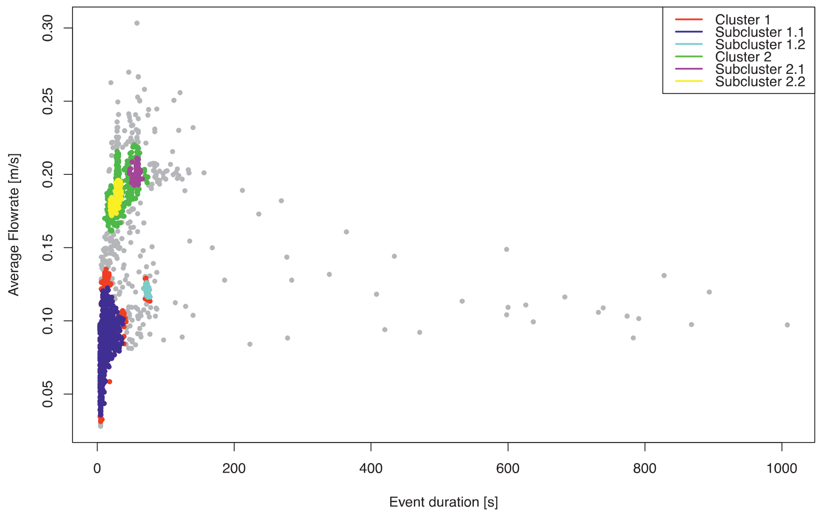

Fig. 3: The duration and average flowrate for the individual clusters

The two main clusters correspond to events with either a high (green) or a low (red) average flow rate. Each of these clusters is then further split up into events with a high and low duration. Especially interesting is the magenta cluster, since nearly all the events in this cluster have a duration of 60 seconds and constant flow during this duration. These events are most likely produced by an appliance. Similarly the cyan cluster has events with duration mostly in the interval of 70-76 seconds and is also most likely produced by an appliance. The events in the both the cyan and magenta cluster also have mostly constant flowrate. The blue and yellow clusters on the other hand are the short events with a wide spread in duration and average flow and therefore most likely events produced by the residents directly.

With the level tree set method, the events with longer duration are not assigned to any clusters, because they occur less often. Amongst these longer events we expect to find the water withdrawals from showers and appliances. These events are more likely to not have a constant flow rate. In this step we only consider the events with duration Therefore, we use a different approach to cluster these events, based on the underlying flowrate time series. We compute the dtw (dynamic time wrap) distance [MR81] between all different pairs of events. This metric tries to make the two time series resemble each other, by stretching and compressing them locally. The distance is computed after stretching by summing the distances of the aligned segments. For the computation we use the implementation in the R-package “dtw” [Gi09] with the Euclidean distance and “symmetric2” step pattern as parameters. Only velocity is used, because the temperature does not have a clear baseline and changes during the day even without any water withdrawal. Having calculated the distance matrix we perform the agglomerative hierarchical clustering [Jo67] with the package “flashClust” [LH12] with the complete linkage method. The resulting dendrogramm is shown in the Fig. 4. There is a clear separation into two main clusters. The cluster on the left side contain the events with more irregular consumption levels and are more likely to be the shower events.

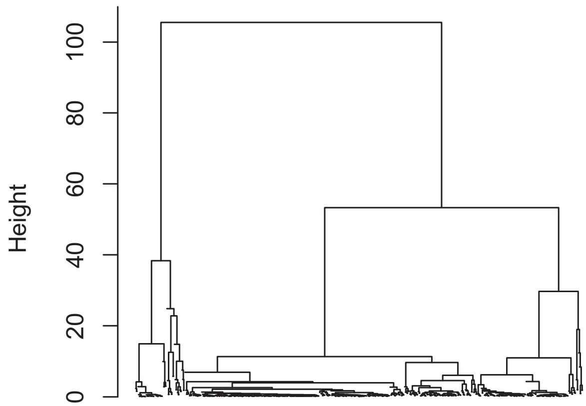

Fig. 4: The result of the hierarchical clustering of the events longer than 30 seconds

4 Conclusions

In this paper we have presented the methodology on disaggregation of the consumption events based on the detailed water consumption data. We can determine three kinds of events: Short events (with average duration less than 40s), regular long events (e.g., appliances), irregular long events (e.g., showers). By categorizing all the consumption events in one of three categories it is possible to provide the residents with a more detailed information on their water consumption (e.g., by showing the cost that occurred due to showers, or the usage of washing machines). The short consumption events are hard to categorize in more details, since they mostly have a constant flowrate.

The presented work can be extended in multiple aspects. A more detailed analysis of the resulting clusters can lead to better insights about the consumption origin. Further in this paper we consider events as separated by zero consumption. But some events are actually multi-part events (e.g., multiple cycles of the washing machine). To recognize such events, we would need first to aggregate nearby events to larger events. During distance calculations with the dtw metric, subpattern matchings (e.g., by calculating the dtw distance with open ends) can then be performed. If there are multiple matches for the sub-patterns, the events can be considered separately. Additionally, the temperature sensor data could be included in the consumption disaggregation to distinguish between the patterns.

References

[At14] Attari, S.Z.: Perceptions of water use. PNAS 111, pp. 5129–5134, 2014.

[Ba08] Bates, B. C.; Kundzewicz, Z. W.; Wu, S.; Palutikof, J. P.: Climate Change and Water - IPCC Technical Paper VI. Intergovernmental Panel of Climate Change (IPCC) Secretariat, Geneva, 2009.

[Br15] Fasy, B. T.; Kim, J.; Lecci, F.; Maria, C.; Rouvreau, V.: The included GUDHI is authored by Clement Maria, Dionysus by Dmitriy Morozov, PHAT by Ulrich Bauer, Michael Kerber and Jan Reininghaus. TDA: Statistical Tools for Topological Data Analysis. R package version 1.4.1, 2015.

[Bu16] Building Energy Data Book. Chapter 8: Water, http://buildingsdatabook.eren.doe.gov/ChapterIntro8.aspx, accessed on 30.05.16

[Do13] Dong, M.; Meira, P. C.; Xu, W.; Chung, C. Y.: Non-intrusive signature extraction for major residential loads. IEEE Transactions on Smart Grid 4/3, pp. 1421-1430, 2013.

[Fi13] Fielding, K.S.; Spinks, A.; Russell, S.; McCrea, R.; Stewart, R.; Gardner, J.: An experimental test of voluntary strategies to promote urban water demand management. Journal of Environmental Management 114, pp. 343–351, 2013.

[Gi09] Giorgino T.: Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. Journal of Statistical Software 31/7, pp. 1-24, 2009.

[Gl03] Gleick, P. H.; Wolff, G.H.; Cushing K. K.: Waste not, want not: the potential for urban water conservation in California. Oakland, CA: Pacific Institute for Studies in Development, Environment, and Security, 2003.

[Ha75] Hartigan, J. A.: Clustering Algorithms, 1975.

[Ha92] Hart, G. W.: Nonintrusive appliance load monitoring, In: Proceedings of the IEEE 80/12, pp. 1870-1891, 1992.

[Jo67] Johnson, S. C.: Hierarchical clustering schemes, Psychometrika 32/3: 241-254, 1967.

[Ke13] Kent, B: P.; Rinaldo, A.; Verstynen, T: DeBaCl: A Python package for interactive DEnsity-BAsed CLustering. arXiv preprint arXiv:1307.8136, 2013.

[LH12] Langfelder, P.; Horvath, S.: Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of Statistical Software 46/11, pp. 1-17, 2012.

[MR81] Myers, C. S.; Rabiner, L. R.: A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected-Word Recognition. Bell System Technical Journal 60/7, pp. 1389-1409, 1981.

[St11] Steinwart, I.: Adaptive Density Level Set Clustering. COLT, 2011.

[Ta15] Tasic, V.; Tiefenbeck, V.; Schöb, S.; Staake, T.: Short-term Spark or Sustained Impact? Investigating the Long-term Effect of Real-time Feedback" ECIS 2015 Research-in-Progress Papers, 2015.

[Ti13] Tiefenbeck, V.; Tasic, V.; Schoeb, S.; Staake, T.: Mechatronics to drive environmental sustainability: Measuring, visualizing and transforming consumer patterns on a large scale, in: Proceedings of the IEEE IECON 2013, Special Section on Energy Informatics, Energy Informatics, 2013.

[To16] Tommorow & Beyond, Water Sense, https://www3.epa.gov/watersense/our_water/tomorrow_beyond.html, accessed on 30.05.16.

[Wi11] Willis, R. M.; Stewart, R. A.; Panuwatwanich, K.; Williams, P. R.; Hollingsworth, A. L.: Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. Journal of environmental management 92/8, 1996-2009, 2011.

[Zo12] Zoha, A.; Gluhak, A.; Imran, M. A.; Rajasegarar, S.: Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. Sensors 12/12, pp. 16838-16866, 2012.