

## SIMULATING APPLIANCE-BASED POWER CONSUMPTION RECORDS FOR ENERGY EFFICIENCY AWARENESS

Mona Ramadan<sup>1\*</sup>, Abdullah Alsalemi<sup>2</sup>, Faycal Bensaali<sup>2</sup>, Abbes Amira<sup>3</sup>, Christos Sardianos<sup>4</sup>, Iraklis Varlamis<sup>4</sup>, George Dimitrakopoulos<sup>4</sup>, Dimosthenis Anagnostopoulos<sup>4</sup>

1 Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA (Corresponding Author)

2 Department of Electrical Engineering, Qatar University, Doha, Qatar

3 Faculty of Computing, Engineering and Media, De Montfort University, Leicester, UK

4 Department of Informatics and Telematics, Harokopio University of Athens, Athens, Greece

### ABSTRACT

The collection and analysis of big data has become a ubiquitous process in numerous fields, including energy efficiency. The combination of smart meters that provide high-frequency readings of consumption and enhanced data analytics have enabled studies to uncover energy usage patterns of households. However, an accurate understanding of consumer behavior requires a more comprehensive approach. Henceforth, we present a novel data simulator that generates appliance-based datasets based on real data. By proper fusion of real smart meter data and periodic energy consumption habits, which are represented by micro-moments, we were able to simulate realistic domestic energy consumption scenarios. Pre-processing the aggregated readings and the application of a k-means clustering algorithm, user actions were detected (switch on/off) per appliance. Based on these records and using a a-priori itemset extraction algorithm we obtained appliance usage patterns. A minute-averaged dataset was simulated for three rooms and nine appliances, monitored for a 47-month period. We managed to extract frequent usage patterns at appliance-level, taking into account contextual parameters such as room occupancy. Moreover, the simulated datasets can be expanded to any size to a variety of energy saving applications such as micro-moment classification and goal-based recommender systems for energy saving.

**Keywords:** Big data, domestic energy usage, energy efficiency, machine learning, micro-moments.

### 1. INTRODUCTION

The advent of big data is a powerful force increasingly influencing decision making of any size, from small societies to whole countries [1], [2]. Being a term applied to datasets whose size is beyond the ability of available tools to process at fast pace, the collection of big data has become a discipline on its own [3], with applications in a multitude of fields [1]. Data repositories such as The University of California Irvine Machine Learning Repository (UCI-MLR) [4], Kaggle [5], and European Union Open Data Portal [6] have become global hubs for research and novel applications, allowing researchers, companies and organizations to share their data with the public to develop smart data solutions. This resulted in a huge expansion of the global datasphere, which is expected to further grow from 33 zettabytes in 2018 to 175 zettabytes by 2025 [7].

In the context of energy, big-volume data is playing an unprecedented role in providing insights on consumption levels and showing both consumers and providers where consumption is excessive and where it is efficient [8], [9]. In fact, the increasing adoption of smart meters, which record detailed and frequent consumption information compared to conventional electricity meters, have provided more granular analytics on usage patterns [10]–[12]. Despite the privacy concerns, the overarching benefits of smart meters

Selection and peer-review under responsibility of the scientific committee of the 11th Int. Conf. on Applied Energy (ICAE2019).

Copyright © 2019 ICAE

outweigh the potential pitfalls, according to experts [13], [14].

Smart meter readings have also assisted in understanding consumer behavior [9], [10], in user profiling, in approximating network load and in predicting short term and long term consumer behavior and forecasting power deals. Data has also been exploited towards understanding socio-demographic information of consumers [15]. Despite the results above, a question still holds: *“is smart meter data adequate for a detailed and comprehensive understanding of household consumption?”*

The overarching goal of gathering energy consumption data is to find practical means of improving efficiency. In a multi-room household that contains a variety of appliances and devices, knowing the overall consumption does not reveal the patterns behind an excessive consumption, nor the specific appliances that led to it. However, by discovering the specific appliances that caused the spike, it would be possible to notify the consumer accordingly and to assist a more informed energy saving decision.

In this work, we argue that there is a potential to improve domestic energy data collection by gathering appliance-based readings as opposed to whole house recordings. For the purpose of our wider study, which aims to develop a goal-based recommender system that will change energy consumption habits of household consumers [16], we need big and detailed energy consumption datasets for households of varying demographics and profiles. The proposed data simulator improves on the way energy consumption data are collected for households, since it generates data at appliance-level and simulates room and house occupancy.

The remainder of the paper is organized as follows. Section 2 lays out the process of generating micro-moment based energy consumption data, and Section 3 discusses integrating real smart meter data into the dataset simulation process. Section 4 depicts current results and relevant applications followed by conclusions in Section 5.

## 2. PROPOSED DATA GENERATOR

### 2.1 Setting up the data generator

The first step in the generation scheme is to specify the set of appliances within the prototype household. Following, power consumption specifications of each appliance is set, which include the minimum, maximum, and standby consumption of the appliance in watt.

After setting the power consumption levels of each appliance in the household, a usage profile for each appliance has to be provided. A usage profile is defined by pairs of timestamps (start-end) that define the time intervals that the appliance is used over the 24 hours of the day. When an appliance has multiple usage patterns (e.g. normal and heavy usage), then multiple sets of timestamp pairs can be provided. The last step is to set the room occupancy rules, which are based on percentages of the daily occupancy of each room. For each room at each time slot a random occupancy flag of 0 or 1 is assigned so that the total ratio of '1' time slots per day reaches the daily occupancy percentage. A data generation system setup for a house of two rooms with a total of four appliances is illustrated in Fig. 1.

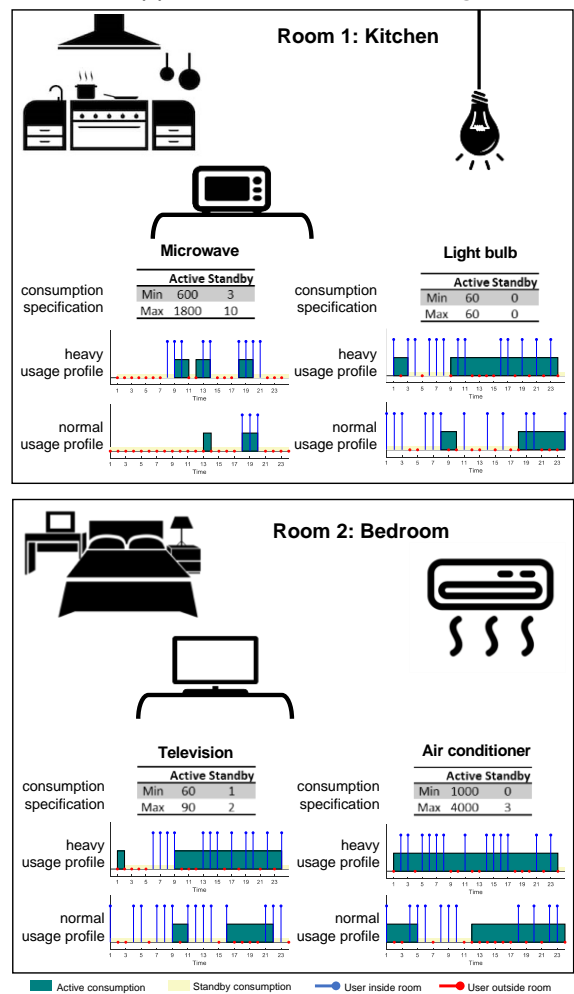


Fig. 1 Data Generation System

### 2.2 Micro-moments for power consumption profiles

Micro-moments could be defined as the isolated incidents particular to a specific spatial-temporal moment [17]. The role of micro-moments may vary depending on the domain they are used in. In the context of energy consumption, micro-moments can be defined

as the occasions when the end-user is ready to perform an action that changes either the house energy consumption or the user status [17]. Similar micro-moments could be merged to formulate a set that could represent the shared essential events for most household appliances as described in Table 1.

Generating power consumption and occupancy records for each appliance and room combination at every time instant is the first step in the construction of a dataset to be used for training energy consumption models. The proposed data generator is designed such that it composes not only power consumption and occupancy values for each appliance, but also assigns one micro-moment label for each time instant [18]. The assignment of micro-moments is based on both the consumption levels and room occupancy while the appliance is in active use.

For the experimental evaluation of this study, the data generator is used to generate data for a household with two rooms described in Fig. 1. Data records of the six appliances are generated for every hour of each day of two consecutive years. Each appliance is assigned a unique numeric appliance ID (0 to 5). Every day of the year is assigned a numeric day stamp (1 to 365), and every hour of the day is assigned a time stamp (1 to 24). The total number of observations in the dataset is 105,120. The generator can produce much bigger datasets that span more years or engage more rooms and appliances.

### 3. DATA-DRIVEN SIMULATION

A successful implementation of this data-driven simulator will allow us to enhance the data generator with more realistic consumption patterns and generate realistic datasets of any size and complexity that combine household demographics with occupancy and consumption habits. We employed a specific power consumption dataset, however, any other dataset that comprises consumption data per appliance, or per group of appliances can be used instead. The Individual Household Electric Power Consumption Data Set available from UCI-MLR [19] contains over two-million measurements of a house in Sceaux, France between December 2006 and November 2010 (47 months).

Table 1 Micro-moments label description.

Numbered label	Label description
0	good consumption
1	switch the appliance on
2	switch the appliance off

3	excessive power consumption
4	consumption of power while outside room

We propose a data simulator that takes as input the power consumption records of the three sub-meters of the French household and converts it to an equivalent dataset that estimates the consumption of each sub-meter for each individual appliance connected to the same sub-meter. The data simulation procedure is split into two main stages. The first stage is a clustering stage that groups the total sub-meter power reading into a number of clusters. The second stage post-processes the output of the clustering step and assigns occupancy and micro-moment labels to the readings. The process of each stage is defined in the next sections.

#### 3.1 Clustering room appliances

The first step for extracting usage patterns from the original measurements data file is to apply a time-series analysis on the consumption information data of each room. More specifically, from the actual energy consumption recorded per minute, we compute the changes between consecutive minutes and between consecutive 5-minutes periods. The first feature allows to isolate minutes where the power consumption increased or decreased significantly due to powering on or off one or more devices. By applying k-means clustering on the different power change values recorded for a room, it is possible to obtain a number of clusters of power change values that can be mapped to specific actions of operating multiple devices.

Based on the consumption values of each device and the devices per room, we map power changes to user actions. Applying the same methodology to the 5-minute changes allowed us to detect power-off actions for devices that go to a low power consumption mode before switching off (e.g. a dishwasher).

#### 3.2 Simulating appliance-level data

The first step is to correct any misjudgments of the clustering algorithm such as pointing to an appliance being switched on when the total sub-meter reading at that record is zero. Once the switching on/off flags are revised, each appliance is assigned a power consumption value (in watt) based on the active and standby consumption ratios for the sub-meter. The active and standby consumption values are available for a wide range of appliances, and the consumption ratios are the ratio of consumption level between different appliances connected to the same sub-meter. For example, when a microwave, a dishwasher and an oven are connected to

the same sub-meter, the active consumption ratio can be 1200 : 1800 : 2400 and the standby ratio can be 7 : 0 : 2. The total sub-meter reading at each time record is then split between all the appliances connected to the same sub-meter based on the ratios. If the total sub-meter reading is above zero while all appliances are shown to be switched off, then the total power is split based on the standby ratio. This step is repeated for each of the three sub-meters within the household and yields individual power records for a total of nine appliances.

The second step is to assign an occupancy flag to the room accommodating each sub-meter. Setting up occupancy rules is based on common sense. The occupancy flag is then chosen to be set to 1 at any time any appliance connected is being switched on or off based on the output of the clustering algorithm.

The final step is to append a micro-moment label to each record of each appliance. The switch on and switch off labels are readily assigned from the output of the clustering algorithm. The remaining labels that need to be assigned to each appliance are: “good use”, “excessive consumption”, and “consumption while outside the room”. An appliance is assigned an “excessive power consumption” flag if it consumes more than 99% of its maximum consumption wattage, or if it is operating on active mode for longer than 99% of the reported maximum operational time. The rest of the recordings are assigned a “good usage” micro-moment.

#### 4. RESULTS AND DISCUSSION

The data generator and simulator are used to generate synthetic data for a period of 24 months and simulated real data for a period of 47 months. In this section, we cast some light on the characteristics and statistics of the produced data. Fig. 2 shows the Micro-moment distribution among the two datasets.

The designed data generator could be used to synthesize data of any size for any user-defined settings. The generator however does not take into consideration differences between weeks, seasons, or even years, which imposes a major limitation. The proposed data simulator yields a very large dataset that contains power consumption profiles for nine different appliances with a total of 18,677,331 observations and 6 features. The features set comprises: appliance ID, recording date, recording time, appliance power consumption, room occupancy flag, and micro-moment label.

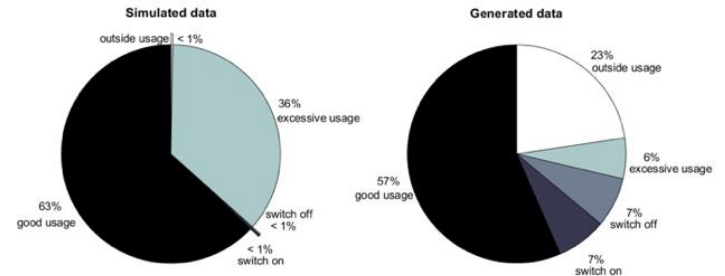


Fig. 2 Micro-moment distribution within the constructed datasets.

We have used the generator produced dataset as means to accurately classify domestic energy consumption profiles by extracting micro-moment from the appliance-based dataset [18]. A supervised, decision-tree based classifier has been trained on the dataset to recognize the patterns in the data to classify unlabeled data records into their corresponding micro-moments labels.

Fig. 3 illustrates how micro-moments are classified during a given day. Feeding the data into the classification algorithm, an ensemble bagged tree classifier, achieved an average testing accuracy of 88%. As part of our future work in this direction, we seek to refine the accuracy further by introducing deep neural networks, which require training on very large datasets and we believe the proposed data simulator will prove useful.

#### 5. CONCLUSIONS

In this paper, we presented a novel data simulator that generates appliance-based datasets based on real data recordings. By combining real smart meter data and periodic energy consumption habits, we simulated realistic domestic energy consumption scenarios with aid from k-means clustering, a-priori extraction algorithm, and the novel use of micro-moments. The simulator produced a minute-averaged dataset for three rooms and nine appliances taking into account usage patterns and room occupancy. Applications include micro-moment classification for highly-detailed consumption profiling and goal-based recommender systems. In this sense, these would be useful for training an energy consumption prediction model or a recommender system for energy related action recommendations. Such energy consumption data generator can be also proven useful in many different scenarios for energy consumption analysis and mining. For example, it could be used to produce large volumes of synthetic consumption data to be used in applications such as deep learning where large datasets are needed for

sufficient algorithm training, or applications of pattern analysis for extracting real-life consumption patterns.

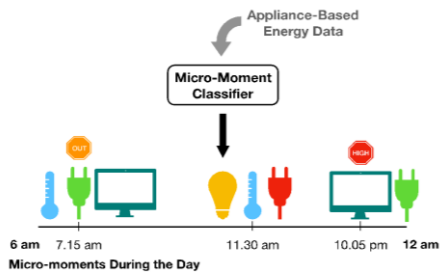


Fig. 3 Classifying micro-moments for a domestic end-user.

#### ACKNOWLEDGEMENT

This paper was made possible by National Priorities Research Program (NPRP) grant No. 10-0130-170288 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

#### REFERENCES

- [1] J. M. Tien, "Big Data: Unleashing information," *J. Syst. Sci. Syst. Eng.*, vol. 22, no. 2, pp. 127–151, Jun. 2013.
- [2] M. M. Gobble, "Big Data: The Next Big Thing in Innovation," *Res.-Technol. Manag.*, vol. 56, no. 1, pp. 64–67, Jan. 2013.
- [3] A. P. Plageras, K. E. Psannis, C. Stergiou, H. Wang, and B. B. Gupta, "Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings," *Future Gener. Comput. Syst.*, vol. 82, pp. 349–357, May 2018.
- [4] "UCI Machine Learning Repository: Data Sets." [Online]. Available: <https://archive.ics.uci.edu/ml/datasets.html>. [Accessed: 07-Mar-2019].
- [5] "Datasets | Kaggle." [Online]. Available: <https://www.kaggle.com/datasets>. [Accessed: 07-Mar-2019].
- [6] "Datasets | European Union Open Data Portal." [Online]. Available: [https://data.europa.eu/euodp/en/data/dataset?sort=views\\_total+desc](https://data.europa.eu/euodp/en/data/dataset?sort=views_total+desc). [Accessed: 07-Mar-2019].
- [7] D. Reinsel, J. Gantz, and J. Rydning, "The Digitization of the World from Edge to Core," 2018.
- [8] H. Jiang, K. Wang, Y. Wang, M. Gao, and Y. Zhang, "Energy big data: A survey," *IEEE Access*, vol. 4, pp. 3844–3861, 2016.
- [9] K. Grolinger, A. L'Heureux, M. A. M. Capretz, and L. Seewald, "Energy Forecasting for Event Venues: Big Data and Prediction Accuracy," *Energy Build.*, vol. 112, pp. 222–233, Jan. 2016.
- [10] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Trans. Smart Grid*, pp. 1–1, 2018.
- [11] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances," in *2012 IEEE International Conference on Pervasive Computing and Communications*, 2012, pp. 190–197.
- [12] D. Alahakoon and X. Yu, "Advanced analytics for harnessing the power of smart meter big data," in *2013 IEEE International Workshop on Intelligent Energy Systems (IWIES)*, 2013, pp. 40–45.
- [13] P. McDaniel and S. McLaughlin, "Security and Privacy Challenges in the Smart Grid," *IEEE Secur. Priv.*, vol. 7, no. 3, pp. 75–77, May 2009.
- [14] E. McKenna, I. Richardson, and M. Thomson, "Smart meter data: Balancing consumer privacy concerns with legitimate applications," *Energy Policy*, vol. 41, pp. 807–814, Feb. 2012.
- [15] K. Zhou and S. Yang, "Understanding household energy consumption behavior: The contribution of energy big data analytics," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 810–819, 2016.
- [16] C. Sardanios *et al.*, "I want to .... change': micro-moment based recommendations can change users' energy habits," in *SmartGreens 2019*, 2019.
- [17] A. Alsalemi, C. Sardanios, F. Bensaali, I. Varlamis, A. Amira, and G. Dimitrakopoulos, "The Role of Micro-Moments: A Survey of Habitual Behavior Change and Recommender Systems for Energy Saving," *IEEE Syst. J.*, 2019.
- [18] A. Alsalemi *et al.*, "Endorsing domestic energy saving behavior using micro-moment classification," *Appl. Energy*, vol. 250, pp. 1302–1311, Sep. 2019.
- [19] Georges HÃ©brail and Alice BÃ©nard, "UCI Machine Learning Repository: Individual household electric power consumption Data Set," 30-Aug-2012. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption#>. [Accessed: 16-Oct-2018].