Energy Disaggregation for Small and Medium Businesses and their Operational Characteristics

Abhinav Srivastava
abhinav@bidgely.com
Bidgely Technologies
Bangalore, Karnataka-India

Paras Tehria
paras@bidgely.com
Bidgely Technologies
Bangalore, Karnataka-India

Basant K. Pandey
basant@bidgely.com
Bidgely Technologies
Bangalore, Karnataka-India

ABSTRACT

Small and Medium Businesses (SMBs) make up a large chunk of the total energy demand, whereas non-intrusive load disaggregation has been dominated by applications in the residential domain. Existing on-site sensor-based approaches for energy disaggregation of SMBs are not scalable and limit energy efficiency analysis, usage behavior study, and kind-of-site classification, which are essential for energy management and grid operations. In this paper, we propose a highly scalable, non-intrusive load disaggregation model for SMBs, which is independent of on-site sensor data.

The energy disaggregation task presented here uses energy data, demographic information, and weather data of SMB's location. The method employs Gaussian Mixture Models, regression models, and DBSCAN to capture appliances' key characteristics. This process can handle usage anomalies, seasonality, and give reliable estimates at the granularity of energy-data sampling rate. Additionally, SMB's working schedule and operational load are estimated using a combination of K-Means and statistical models. Operational load signatures and working schedule and appliance characteristics can further be used for energy management, efficiency analysis, and site classification.

The model has been evaluated using anonymized energy data of different types of commercial spaces such as office sites, restaurants, individual shops, and shopping complexes from the USA (North Carolina, South Carolina, Indiana, and Florida) with total monthly consumption ranging from 800 kWh to 120,000 kWh. Results indicate reliable disaggregation of load into the heating, cooling, and baseload categories along with an estimate of operational hours and operational load at the energy-data sampling rate.

CCS CONCEPTS

• **Computing methodologies** → *Machine learning*; **Unsupervised learning**; *Feature selection.*

KEYWORDS

Energy disaggregation, Small and medium businesses, Commercial sites, Unsupervised machine learning

NILM'20, November 18, 2020, Virtual Event, Japan

ACM ISBN 978-1-4503-8191-8/20/11…$15.00
https://doi.org/10.1145/3427771.3427854

ACM Reference Format:
Abhinav Srivastava, Paras Tehria, and Basant K. Pandey. 2020. Energy Disaggregation for Small and Medium Businesses and their Operational Characteristics. In *proceedings of The 5th International Workshop on Non-Intrusive Load Monitoring (NILM'20), Nov 18, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3427771.3427854

1 INTRODUCTION

Non-intrusive Load Monitoring (NILM) is the method of getting appliance-level consumption from total energy usage without using physical sub-meters. Getting appliance-level disaggregation is very useful as it helps end-users understand their energy usage behavior and motivates them to save energy. Due to its merits, NILM has gained a lot of attention from researchers worldwide, and the number of research articles in this field is increasing[2, 5, 7]. However, most of the research work in this field is limited to residential buildings.

In 1996, LK Norford and S.B. Leeb published a paper, where they used steady-state and transient load-detection algorithms to disaggregate the electric load of commercial buildings [6]. The results of this research were relatively inaccurate because of the complexity of appliances in commercial buildings and the lack of adequate training data.

In recent times, there have been several novel approaches used for NILM of commercial sites. A significant issue that persists is the lack of reliable, publicly available datasets. Owing to this, most of the approaches rely on a small number of data points. Ji et al. used a Fourier series model for disaggregating HVAC load in industrial sites in their research paper[4]. They used ground truth data of very few locations. Although promising, this approach has not been proven to work on a diverse and more extensive set of commercial sites. Similar problems exist in other research work done using a small datasets[1, 3].

In general, in the absence of appliance-level sub-meter data, disaggregation for SMBs is done using statistical rules and heuristics, which results in highly inaccurate estimates. There is a need to come up with Non-Intrusive-Load-Disaggregation algorithms for SMBs, which are accurate and personalized for a specific site, hence facilitating accurate usage insights. In this paper, we have tried to evade the tagged data limitation problem and came up with a sequence of unsupervised algorithms for energy disaggregation of SMBs. The algorithm focuses on characterizing an SMB site based on heating, cooling, always-on, anomalous, and operational consumptions. These appliance estimates are reliable at the granularity of the energy-data sampling rate. Additionally, in the process of disaggregation, characteristic features are derived that facilitate actionable insights for SMBs.

2 SMB BACKGROUND

SMBs encompass a wide range of energy consumption levels depending upon their size, type, and location. Also, within an SMB, consumption levels may vary across months. One of the primary features of an SMB is its characteristic working hours, which usually remain similar throughout the year, with minor seasonal/daily variations. During working hours, there are a few operation-related appliances, whose usage generally remains the same, or is a function of only the hour of the day. For example, in a small IT business, consumption categories like lighting, hub units, refrigeration, etc. run continuously during the open hours. Consumption categories like desktop units depend on the time of the day as per the employee schedule. We call this component of energy usage as Operational load (E_{op}). In general, during close hours, consumption levels are low, but there might be some anomalies or day-specific variations that require special attention.

Heating Ventilation and Air Conditioning (HVAC) systems operate seasonally, i.e., cooling occurs during summers, and heating occurs during winter. Most HVAC usage is limited to working hours. Still, a spillover is possible in closed hours due to inefficient usage behavior.

3 DISAGGREGATION PIPELINE

3.1 Datasets

In order to develop a method for disaggregation, we considered 1022 SMBs with Advanced Metering Infrastructure (AMI). These SMBs are picked up from North Carolina, South Carolina, Indiana, and Florida regions of the USA, and have at least one year of data. The sampling interval of raw energy usage data (E_{raw}) is at three discrete levels, 15-min, 30-min, and 60-min. As there are appliances that show seasonal variations, open-source weather information of the SMB's location is also included in the process to capture such variations.

3.2 Data Pre-processing

Unsupervised algorithms used in building the disaggregation process are sensitive to data outliers. In order to handle such issues, we pre-processed consumption and temperature data by handling anomalous consumption, anomalous temperature points, and imputing missing consumption or temperature data.

3.3 Working Hour Estimation

Working hours estimation is carried out in two steps. First, we estimate a general trend of working hours, which remains consistent with slight variations in the local chunk of days. Second, we capture deviations from the general trend caused by various reasons such as festivals, day of the week, vacation season, extended or shortened working hours.

In general, working hours have significantly higher energy consumption levels in comparison to non-working hours. During non-working hours, the major contribution is from always-on appliances. On a working day, energy demand picks up at about opening time, but the day's always-on component remains the same, creating an energy contrast. For calculating working hours, we leverage this contrast in consumption. Removing the daily always-on energy consumption (E_{day-ao}) further amplifies this contrast, which helps in making accurate working hour estimates.

$$E_{contrast} = E_{raw} - E_{day-ao} \tag{1}$$

3.3.1 General Trend. A differential increase in energy (d) is obtained from $E_{contrast}$ for each energy-data point. This energy differential when arranged in 2-dimensions :

- hour of the day (h)
- all available days (D)

creates signed edges for open and close time. For a chunk of days (n) these edges are detected at the hours h having most likely potential (P_h) of open or close edges.

$$d_h = E_{contrast_{(h)}} - E_{contrast_{(h-1)}} \tag{2}$$

$$P_h = \sum_{D=1}^{n} d_h \tag{3}$$

The hour with the lowest potential value becomes general opening time, while the hour carrying the highest potential value becomes general closing time.

3.3.2 Day Specific Variations. The general trend of working hours is a characteristic feature for a site, but it misses upon day-specific variations. In order to capture this, $E_{contrast}$ is clustered into two components using K-Means clustering. This method may miss upon general trend but captures well the day-specific variations in working hours.

The general trend and day-specific variations of working hours are evaluated every month, and their union is taken to accurately classify each energy-data sampling points into open or close hours.

3.4 Disaggregation Approach

We have defined the baseload as a consumption component that runs throughout the day and also has seasonal consistency. HVAC is purely seasonal in nature with two sub-categories based on the duration of operation, on-demand and always running. To get a fair estimate of HVAC, the baseload is estimated and removed from E_{raw} before proceeding with HVAC estimates. In working hours, apart from baseload and HVAC, another major consumption component is a perennial operational load, which varies as a function of the hour of the day. Consumption apart from HVAC and baseload during working hours, is used as a basis to estimate operational load, which also acts as a characteristic feature for SMB related analytics.

3.4.1 Baseload Estimate. Appliances contributing to the baseload run year-round. We have estimated the baseload consumption ($E_{baseload}$) by extracting seasonal consistencies at the low energy spectrum of E_{raw}. Before going for HVAC estimates, we remove the $E_{baseload}$ from E_{raw} to get E_{r-b}. Where,

$$E_{r-b} = E_{raw} - E_{baseload} \tag{4}$$

3.4.2 HVAC Estimate. It is general practice to model individual appliances in NILM, but getting accurate HVAC estimates from *appliance-based* modeling for SMBs is difficult, as a wide range of combinations are possible between HVAC demand, number of occupants and number of HVAC appliances. In the absence of appliance specific sensor data, it is even more challenging. Each of the unique combinations mentioned above can be identified as a unique HVAC *usage-behavior*.

We have taken a *usage-behavior* based modeling approach to get HVAC estimates instead of *appliance-based* modeling. All the *usage-behaviors* are identified and segregated, and then *sub usage-behavior* models are built to get HVAC estimates. We have taken a seasonal, multiple mode modeling approach using E_{r-b} to segregate desired *usage-behaviors*. A gaussian mixture model (GMM) is used to separate HVAC *usage-behavior* mode (*HVAC-mode*) based on distribution of E_{r-b} at data-sample level. The mode-characteristic generated is captured using the mean and the standard deviation of each mode. These *HVAC-modes* have inherent seasonality associated with them. HVAC appliances consume more energy when operated in harsher ambient conditions in comparison to milder conditions. Using a temperature derived feature F_t, a regression model is learnt with E_{r-b} as dependent variable and F_t as independent variable to capture the seasonality within each of the *HVAC-modes*. An *HVAC-mode* may contain multiple regression-trends between E_{r-b} and F_t. These trends are separated using DBSCAN. We identify each trend as HVAC *sub usage-behaviors* and they mimic *appliance-based* modeling. Evidently E_{r-b} carries two kind of regression functional relation with F_t. A linear relation is the most common relation in SMBs, but a square root trend might also be present if one or more of HVAC appliances falling in a *sub usage-behavior* are heavily loaded and tend to saturate.

$$Linear : E_{r-b} = m * F_t + intercept \tag{5}$$

$$Square - root : E_{r-b} = m * F_t^{1/2} + intercept \tag{6}$$

In equation 5 and 8, HVAC coefficient (m) and *intercept* learnt carries characteristic information about each HVAC *sub usage-behavior*. These parameters are learnt for heating and cooling separately. Figure 1 shows the distribution of heating coefficients of SMBs mentioned in sub-section 3.1, separated in three energy consumption ranges.

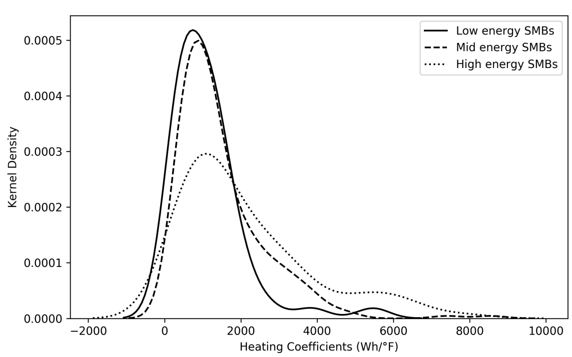

Figure 1: Heating coefficients of SMBs

Larger the coefficient, higher the consumption increase of the HVAC appliances in $subusage-behavior$ for a unit rise in F_t. Hence, the high coefficient indicates that HVAC usage could be inefficient.

HVAC estimates(E_{hvac}) are made at data-sampling level using coefficient and F_t value at each data sample. Such estimates are made for each HVAC *sub usage-behavior* identified through GMM and DBSCAN separations. The estimates are merged from all *sub usage-behaviors* to give net HVAC consumption for the entire year at the energy-data sampling level.

3.4.3 Operational Load Estimate. Operational load estimates are, by definition, valid only for working hours of SMB. After the baseload estimates and the HVAC estimates are subtracted from E_{raw}, the resultant residual energy($E_{residual}$) carries information about the operational load.

$$E_{residual} = E_{raw} - E_{baseload} - E_{hvac} \tag{7}$$

The operational load(E_{op}) is a function of hour of the day. At each hour of the working hours, the E_{op} is estimated using a generalized linear model (GLM) on $E_{residual}$, in a month. The monthly estimates takes care of seasonal variation in E_{op}, if any. The E_{op} estimated is assigned to respective working hour, in every month. The average operational load signature (a 24-hour vector) becomes a characteristic for an SMB, which can be used as a feature for SMB classification.

4 RESULTS AND DISCUSSION

Disaggregation methodology discussed in the previous section generates data-sampling level energy estimates of HVAC consumption, always-on consumption, and the operational load. The accuracy of HVAC estimation is evaluated against ground truth data from North Carolina (NC). We also analyze the TOU estimates and SMB characteristics on the user-set mentioned in sub-section 3.1 to demonstrate the overall validity of the proposed approach.

4.1 Evaluation: NC Ground Truth Data

The ground truth includes data from 3 offices and 5 shops. Accuracy of the HVAC estimation has been evaluated at monthly level as shown in Table 1. The metric used is mean absolute percentage accuracy, calculated as:

$$Accuracy = 100 * \frac{\sum_{i=1}^{n}(1 - \frac{|Estimated_i - True_i|}{True_i})}{n} \tag{8}$$

Table 1: Estimation accuracy on SMBs from NC

SMB type	SMB count	Estimation Accuracy %	
		Cooling	Heating
Offices	3	88.3 - 92.1	81.0 - 89.5
Shops	5	85.5 - 90.3	80.8 - 88.4

4.2 Appliance: Time of Usage (TOU)

Figure 2 shows the appliance TOU of an SMB from Florida. The shaded portion shows the working hours. We can infer from the figure that HVAC consumption remains almost the same in working hours with minor hour-specific variations, which indicates that the HVAC system is running at its maximum capacity. This information can be extracted qualitatively using the heating characteristic

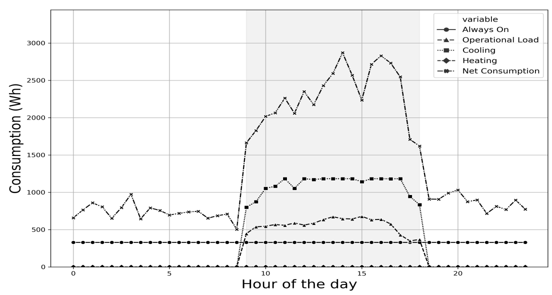

Figure 2: Appliance TOU of an SMB from Florida.

coefficient from regression and mode amplitude. Additionally, the operational load is observed to increase post mid-hours of open duration. Unaccounted fluctuations may have risen from unknown on-demand consumption on that day.

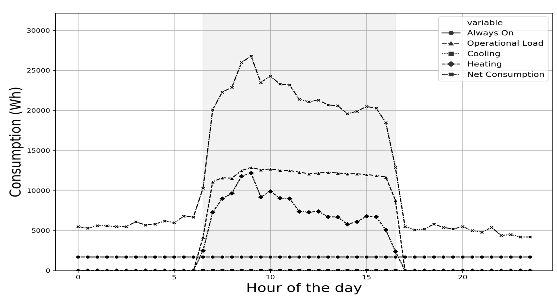

Figure 3: Appliance TOU of an SMB from North Carolina.

Figure 3 shows the TOU of an SMB from North Carolina. Shaded portion shows the working hours. Unlike the previous illustration, the HVAC system is causing fluctuations in energy consumption during working hours, whereas the operational load remains almost flat.

There is a substantial difference in the operational load between the two SMBs. For the SMB in Figure 2, it varies around 600 Wh. In comparison, for the SMB in Figure 3, it hovers at about 12,500 Wh. The difference in operational load can originate from a different set of operation related appliances, working hours, and other characteristics and hence can be a feature for SMB type classification.

4.3 SMB Characteristics

In addition to appliance TOU, the disaggregation approach described in the paper has given us characteristic features like HVAC *usage − behavior* modes(mu, sigma), HVAC coefficients, working-hours, and operational load signature. These characteristic features give vital insights that, in conjunction with additional meta-information, can be utilized for inter-SMB comparison and classification.

For instance, Figure 4 shows the distribution of cooling coefficients for SMBs mentioned in sub-section 3.1. As described in the sub-section 3.4.2, the inefficiency of HVAC (in this case, cooling)

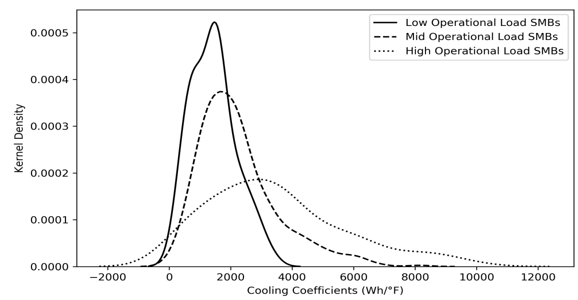

Figure 4: Cooling coefficients of SMBs with similar characteristic operational load

usage increases as we move towards the far right of the coefficient spectrum.

SMBs with larger operational loads show a higher cooling coefficient, which indicates either a larger floor area or a higher number of occupants. The ratio of operational loads within similar buckets has the potential of generating a feature for SMB classification.

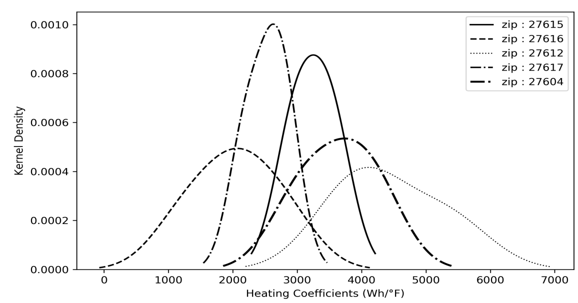

Figure 5: Comparison of cooling coefficients of SMBs from Raleigh city

Figure 5 shows the comparison of heating coefficient distribution for zip-codes in Raleigh city in North Carolina. A clear separation can be seen in mode placement, which indicates that there may be regions in the city that either have more efficient heating systems or may have smaller SMB sites.

5 CONCLUSION

The unsupervised disaggregation approach developed can explain a large portion of SMB energy consumption, without any reliance on appliance-level sub-meters. This approach gives us appliance-specific time-of-usage information at the energy-data sampling rate and provides characteristics that carry essential information that can be leveraged for applications in energy efficiency, usage behavior, and site classification domains. This method is scalable and can be applied to any geography which has SMBs with AMI energy data.

ACKNOWLEDGMENTS

We would like to acknowledge the support and encouragement from the executive team at Bidgely that made this paper possible. We would also like to note the efforts of the complete data science team at Bidgely for their numerous contributions to the project and the paper.

REFERENCES

[1] Shubham Bansal and Mischa Schmidt. 2017. Energy Disaggregation Methods for Commercial Buildingsusing Smart Meter and Operational data. In *31st AAAI Conference on Artificial Intelligence (AAAI-17, San Francisco, 4–9 February 2017*. AI Access Foundation, 325–329.

[2] S. Drenker and A. Kader. 1999. Nonintrusive monitoring of electric loads. *IEEE Computer Applications in Power* 12, 4 (1999), 47–51. https://doi.org/10.1109/67.795138

[3] Mohammad A Hossain, Arash Khalilnejad, Rachel A Swanson, Jack Mousseau, Ethan M Pickering, Roger H French, and Alexis R Abramson. 2017. Unsupervised Non-Intrusive Energy Disaggregation for Commercial Buildings. (2017).

[4] Ying Ji, Peng Xu, and Yunyang Ye. 2015. HVAC terminal hourly end-use disaggregation in commercial buildings with Fourier series model. *Energy and Buildings* 97 (June 2015), 33–46. https://doi.org/10.1016/j.enbuild.2015.03.048

[5] C. Laughman, Kwangduk Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong. 2003. Power signature analysis. *IEEE Power and Energy Magazine* 1, 2 (2003), 56–63.

[6] Leslie K. Norford and Steven B. Leeb. 1996. Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms. *Energy and Buildings* 24, 1 (Jan. 1996), 51–64. https://doi.org/10.1016/0378-7788(95)00958-2

[7] Hannu Pihala. 1998. *Non-intrusive appliance load monitoring system based on a modern kWh-meter.* Technical Research Centre of Finland.