

LEAD1.0: A Large-scale Annotated Dataset for Energy Anomaly Detection in Commercial Buildings

Manoj Gulati
manojg@smu.edu.sg
Singapore Management University
Singapore

Pandarasamy Arjunan
samy@bears-berkeley.sg
Berkeley Education Alliance for Research in Singapore
(BEARS) Limited, Singapore

ABSTRACT

Modern buildings are densely equipped with smart energy meters, which periodically generate a massive amount of time-series data yielding a few million data points every day. This data can be leveraged to discover the underlying load and infer their energy consumption patterns, inter-dependencies on environmental factors, and the building’s operational properties. Furthermore, it allows us to simultaneously identify anomalies present in the electricity consumption profiles, which is a big step towards saving energy and achieving global sustainability. However, to date, the lack of large-scale annotated energy consumption datasets hinders the ongoing research in anomaly detection. We contribute to this effort by releasing a carefully annotated version of a publicly available *ASHRAE Great Energy Predictor III* data set containing 1,413 smart electricity meter time series spanning over one year. In addition, we benchmark the performance of eight state-of-the-art anomaly detection methods on our dataset and compare their performance.

CCS CONCEPTS

• **Applied computing** → *Physical sciences and engineering*; • **Computer systems organization** → *Sensor networks*; • **Computing methodologies** → *Anomaly detection*.

KEYWORDS

Smart buildings, smart meters, time-series analysis, outlier detection, anomaly detection, and machine learning.

ACM Reference Format:

Manoj Gulati and Pandarasamy Arjunan. 2022. LEAD1.0: A Large-scale Annotated Dataset for Energy Anomaly Detection in Commercial Buildings. In *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22)*, June 28–July 1, 2022, Virtual Event, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3538637.3539761>

1 INTRODUCTION

Buildings are one of the largest energy consumers, accounting for approximately 40% of the total energy usage in the world [12]. It is estimated that 20% of the total energy consumed gets wasted within buildings [11]. Further, the energy demand of the buildings is increasing continuously and will rise by 28% by 2040 [5]. Hence there is a pressing need to reduce energy wastage to lower the energy footprint and cost of buildings’ utilities. The most common

causes of energy wastage are equipment failure, aging, misconfiguration, and human-related errors. This *wasteful use of energy* which is mostly quantified as *unusual energy usage* can be identified and prevented using a data-driven analytical technique known as *anomaly detection* [7, 10].

More recently, the fast-paced proliferation of smart meters has led to a boost in research focused on leveraging smart meter data for anomaly detection in buildings [1]. This surge in anomaly detection research is due to (i) the realization of the significance of the contribution of buildings towards gross energy consumption and (ii) the benefits it can garner in terms of long-term energy sustainability in buildings. Also, it is the most efficient mode of sensing and detection due to ease of installation, monitoring, and scalability. Anomaly detection in building energy consumption is the process of identifying unusual energy usage events that lead to energy waste. Such abnormal use differs significantly from the normal energy usage patterns. The proliferation of advanced metering infrastructure (AMI) [6], along with improved computationally intelligent methods, makes it possible to develop automated anomaly detection techniques [7]. These techniques track the building’s energy consumption patterns from the aggregate smart meter data, identify anomalous events of energy consumption and report them to building managers for further action.

A vast literature on anomaly detection techniques exists for time series in various domains [4]. In [7], the authors describe some of the practical challenges in detecting anomalies in energy consumption data and reviewed different approaches namely (a) statistics-based rule sets, (b) unsupervised, and (c) supervised. While anomaly detection in building energy consumption is an active area of research, several challenges exist and hampers its acceptance in the real world as a method to screen and optimize energy utilization [7]. This include (a) difficulty in collection and assignment of anomaly labels, (b) lack of annotated public dataset for anomaly detection research, furthermore, (c) due to a lack of annotated dataset, people have widely used unsupervised methods, which lead to a higher rate of false positives. Consequently, building operators have to go through all identified anomalies and filter out the genuine ones to take further actions, which is a tedious and time-consuming task. Due to this limitation, existing studies have evaluated the models on a few buildings, making it difficult to estimate the true energy-saving potential of the models.

To mitigate these limitations, we annotate and release **LEAD1.0 – a Large-scale Energy Anomaly Detection**¹ dataset consisting of 1,413 smart electricity meter data spanning over an year. To the best of our knowledge, this dataset is so far the largest for

e-Energy '22, June 28–July 1, 2022, Virtual Event, USA

© 2022 Association for Computing Machinery.

This is the author’s version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22)*, June 28–July 1, 2022, Virtual Event, USA, <https://doi.org/10.1145/3538637.3539761>.

¹<https://github.com/samy101/lead-dataset>

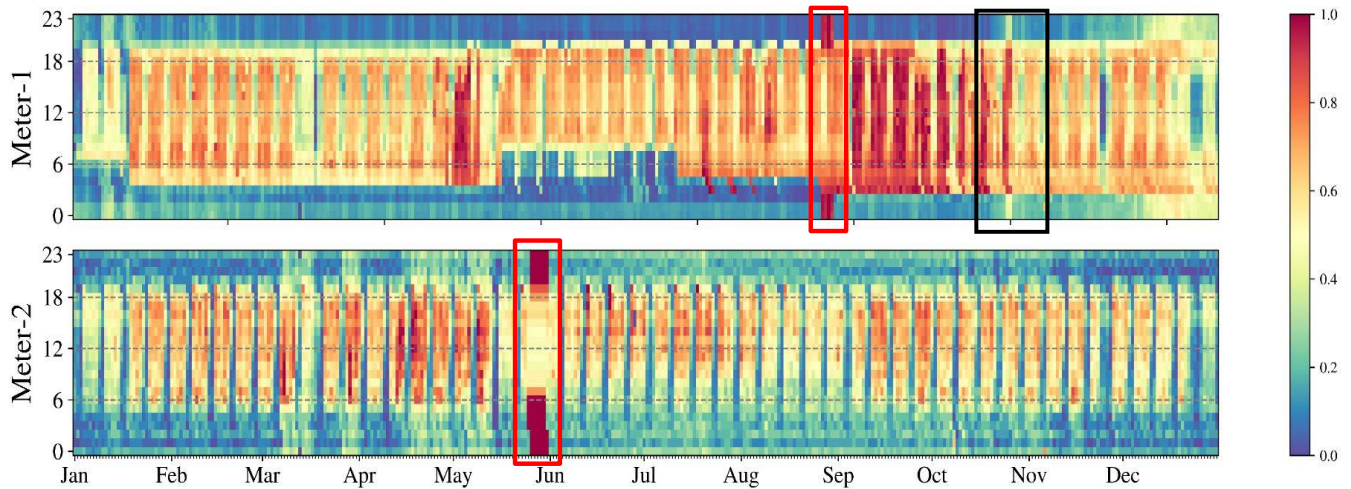


Figure 1: The hourly electricity usage patterns (normalized) of two buildings with examples for point and sequential anomalies. It shows, (a) daily and weekly power usage cycles, (b) examples of single point anomaly (marked in black), and (c) examples of sequence anomaly (marked in red). The X-axis denotes the day of the year, and Y-axis is the hour of the day. The colors represent the electricity usage at that hour, with blue being lower and red being higher usage.

energy anomaly detection in the public domain. We also benchmark the performance of several state-of-the-art approaches and release the code-base of anomaly annotation tool in open-source for community use.

2 DATA SET PREPARATION

We leveraged the dataset used in the Great Energy Predictor III competition² conducted in 2019 on the Kaggle platform [8]. This dataset includes one year of hourly meter readings from 1,636 non-residential buildings collected from 16 different sites worldwide [9]. Also, it contains building meta-data like *square_feet*, *year_built*, and *floor_count* to describe the structure of the building (specified by the *building_id*). Furthermore, it is accompanied by various weather parameters to help model the buildings' energy usage better.

This dataset had measurements taken from four different energy meter types (electricity, chilled water, steam, and hot-water). For the task of anomaly detection, we exercised hourly meter readings data from 1,413 electricity meters covering 16 different building types, such as office, monitored for one year. Please note that in the original ASHRAE dataset, there were 1,636 buildings (not meters). Each building had different energy meter types such as electricity, hot-water, etc. In this paper, we have focused only on the electricity meters (1,413) and in future we plan to annotate other meter types as well. The top five winners of this competition have annotated some outliers that they excluded for model development; however, these annotations were not comprehensive and were missing class labels for different types of anomalies.

In our expedition, we annotated this dataset with (a) point anomalies and (b) sequential or collective anomalies:

- (1) **Point anomaly:** A point anomaly occurs when an individual point can be considered as an anomaly compared to the rest of the energy consumption data [2]. Also, a point anomaly can be explained as an energy consumption instance that appears unusual when compared to the overall/whole time series (global) or compared to its neighboring points (local). It occurs once at any time and does not repeat. Figure 1 shows the example of a single point anomaly (a single day within the black box) in Meter-1.
- (2) **Sequential or collective anomaly:** A sequential anomaly refers to a consecutive set of energy consumption events whose joint behavior is unusual. It may occur once or repeatedly at regular intervals [1]. It can also be explained as a collection of related data instances which are anomalous with respect to the entire data set. The individual data instances in a collective anomaly may not be anomalies by themselves, but their occurrence together as a collection is anomalous [3]. Sequential anomalies can also be local or global. Figure 1 shows examples of sequence anomalies (red boxes) in Meter-1 and 2 as some power-intensive appliances were turned-on during non-working hours for a few days in a row.

We have developed a web-based tool to annotate every point in the electricity meter time series. This process involved manually examining approximately 12 million data points in total, with each inspection window having 8,784 data points on an average for each electricity meter. The original ASHRAE dataset contains hourly readings for the entire year of 2016. Since 2016 is a leap year, there were 8,784 (366x24) samples for each meter. Each manual inspection (over 24 hour window span) took somewhere from three to five minutes, collectively accounting for approximately 100 man-hours invested for this strenuous annotation exercise. Please note, that we have carefully annotated with an hourly data span, and not full

²<https://www.kaggle.com/c/ashrae-energy-prediction/>

days. The current benchmarking study (discussed in Section-3) is focused on detecting whether a day has any anomaly or not, inline with some previous studies found in the literature. However, it is quite possible to pinpoint the anomalies in our dataset using a different experimental setup with the same PyOD models.

We followed a fixed protocol for annotating each time window (as discussed below). For each meter’s time-series plot, we zoom in and look for a weekly or daily pattern in the time series of meter readings. (i) If the plot follows a weekly or daily pattern, then we look for the disturbance in this pattern. And that disturbance is marked as an anomaly. (ii) If the plot does not have any suitable pattern, then we look for the days having higher or lower energy consumption than the usual. If we find that there is a large difference between that day’s energy consumption and its nearby days, then we mark that as an anomaly. It is hard to follow the same set of rules for annotating anomalies in all the buildings because each building has a different definition of anomaly but above mentioned steps are the most common steps that we applied while doing annotation. Also, the tool that we have created provides us a better and efficient way to annotate. We have released this web-tool along with this dataset for the public use. Please note, these anomalies were defined based on literature and after going through each raw time series and also enforced a verification process as part of our anomaly annotation protocol. Precisely, we had 199,640 (1.66%) anomalous instances present across 12,060,910 data points logged by 1,413 energy meters (excluding the missing data points which would make the total count of 12,411,792 data points i.e. 1,413 meters x 8784 [366days x 24hrs]).

3 BENCHMARKING

3.1 Data pre-processing

We first create new categorical features based on the timestep - hour, day, weekday, and month. Then, we normalize all the available features using the z-score normalization technique. We don’t directly use the meter-reading data but calculate its natural logarithm. We have used $\log(p+1)$ for data-normalization, as this is a standard way to scale the target before fitting the model and use it for prediction purposes. Post data-normalization, we group the data available for each building and apply a sliding window across the available timesteps. In this paper, we have experimented with 24-hour windows with zero overlaps, so that models try to learn daily energy patterns for each building.

3.2 Baseline Models

We have considered a variety of statistics-based methods as well as machine learning-based methods for benchmarking anomaly detection on our dataset. Conventionally, clustering-based algorithms are already being employed in various applications where anomaly detection has been studied. In our work, we evaluate some of these techniques, which include (a) Cluster-based Local Outlier Factor (CBLOF), (b) Feature Bagging, (c) Histogram-base Outlier Detection (HBOS), (d) Isolation Forest, (e) K-Nearest Neighbors (KNN), (f) Local Outlier Factor (LOF), (g) Minimum Covariance Determinant (MCD), and (h) One-class SVM (OCSVM). All these models were

implemented in Python with the help of the PyOD library³ and executed on Google Colab platform.

3.3 Evaluation Metrics

In this work, we have used Precision, Recall, and F1 score to have a fair understanding of the distribution of false positives and false negatives while determining anomalous profiles. The formulas are as follows -

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (1)$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (2)$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

3.4 Experimental setup

Firstly, we separate the anomalous and the non-anomalous data. The non-anomalous samples were split into train(80%), validation(10%) and test(10%) sets, and the anomalous samples were split into train(10%), validation(20%) and test(70%) sets. Both the non-anomalous and anomalous sets were combined to make the final training, validation, and test tests. The model is trained on the training sets and then validated using validation set. The threshold calculated based on the validation sets is then used on the test sets to measure the model performance.

3.5 Model comparison

Please note that we are not interested to specifically pinpoint the time point of the anomalous activity. However, in this work, we focus on identifying whether the entire 24-hour sequence is anomalous or not. We compute F1-Score, Precision and Recall (as mentioned previously) on our annotated dataset. These results are presented in Table 1. K-Nearest Neighbors and Minimum Covariance Determinant yield the highest precision i.e. 0.902 and 0.901 respectively. In terms of recall, KNN stands out by yielding 0.284 recall followed by Local Outlier Factor whose recall is around 0.281. In terms of F1-score, obviously KNN outperforms all the models by yielding an F1-score of 0.431, followed by Local Outlier Factor whose score is around 0.426. We acknowledge that we have only tried to conduct a limited benchmarking exercise here, and the list of models that we have experimented with is no where an exhaustive list and definitely there are plenty of other models which can be tried out on this data.

4 DISCUSSION AND CONCLUSIONS

Data-driven energy sustainability in buildings is considered to be a major leap in reducing carbon footprint and asserting the pace of the ongoing climate change. As Buildings are one of the significant energy consumers, they are a strong contender for energy optimisation with a long-term goal of energy sustainability. Among various techniques proposed to emend energy consumption in the buildings, data-driven energy monitoring and energy anomaly detection serve the highest priority. This preference is due to inherently simple deployments of energy meters, absence of cumbersome retrofits and

³<https://github.com/yzhao062/pyod>

Table 1: Comparison of F1-score, Precision and Recall of all anomaly detection models.

Model	F1-score	Precision	Recall
Cluster-based Local Outlier Factor (CBLOF)	0.425	0.900	0.277
Feature Bagging	0.424	0.899	0.279
K-Nearest Neighbors (KNN)	0.431	0.902	0.284
Histogram-base Outlier Detection (HBOS)	0.397	0.896	0.258
Isolation Forest	0.413	0.895	0.270
One-class SVM (OCSVM)	0.421	0.899	0.276
Local Outlier Factor (LOF)	0.426	0.900	0.281
Minimum Covariance Determinant (MCD)	0.422	0.901	0.276

up-gradation of infrastructure needed for these techniques. Anomaly detection caters to this broader goal by identifying avoidable electricity usage and reporting it to the stakeholders.

Regardless, the data processing techniques are still not streamlined and standardized, especially the data annotation process. Due to the practical difficulties in collecting accurate anomaly labels along with meter data, manual data labeling is still the default mode used to annotate the meter readings as a post-processing step. We

also acknowledge Bhavuk Singhal and Mudit Dhawan for their contribution towards generating annotations and assistance in establishing the benchmarks.

REFERENCES

- [1] Pandarasamy Arjunan, Harshad D Khadilkar, Tanuja Ganu, Zainul M Charbiwala, Amarjeet Singh, and Pushpendra Singh. 2015. Multi-user energy consumption monitoring and anomaly detection with partial context information.