Heuristics-Based Detection of Abnormal Energy Consumption

Ankur Sial[1], Amarjeet Singh[1], Aniket Mahanti[2], and Mingwei Gong[3](✉)

[1] IIIT-Delhi, Delhi, India
[2] University of Auckland, Auckland, New Zealand
[3] Mount Royal University, Calgary, Canada
mgong@mtroyal.ca

Abstract. This paper presents two methods for detecting abnormal electricity consumption by utilizing usage patterns in the vicinity. The methods use contextual and factual information including, energy consumption patterns, nature of supply and category of day to logically group meters and find abnormalities. Using heuristics proposed in the paper, data collected from fifty smart meters deployed inside hostels of IIIT-Delhi were investigated for abnormal electricity consumption. Multiple abnormalities were found and their causes were verified after discussion with campus administrators. Our results show that the proposed heuristics successfully found abnormal energy consumption behavior. Therefore, these methods could be used for real-time abnormality detection. This will result in reducing operating costs by automatically detecting and reporting abnormalities without human intervention.

1 Introduction

Many countries across the globe have started to realize the importance of energy efficiency. Efficient energy usage habits reduce overall expenditure on energy. Energy conservation reduces energy wastage, contributing towards a sustainable development.

Smart meters are the next generation of electricity meters. They help in better energy accounting. Using smart meters, users can make informed decisions for reducing energy usage. Smart meters provide fine grained data that can be used for monitoring, recognition, and profiling of appliances using device signatures.

For efficient energy usage, abnormal behavior should be identified. While developing prediction techniques and benchmarking mechanisms one has to remove anomalous data. Therefore, new methods for abnormal energy consumption detection should be devised to help identify abnormal system behavior. Visualization of energy consumption using smart meters helps in finding abnormal behavior and reduce energy costs. Using smart meters, a better demand response system can be created through pattern analysis.

The main contribution of this paper is to apply two heuristics for detecting abnormal energy consumption detection in a residential campus. Using smart

P. H. J. Chong et al. (Eds.): SmartGIFT 2018, LNICST 245, pp. 21–31, 2018.
https://doi.org/10.1007/978-3-319-94965-9_3

meters we track energy consumption for hostel residents in the campus. To the best of our knowledge, previous research analyzed energy consumption at a house or a building level. We are the first to detect abnormal energy consumption for a group of meters having same context inside hostels in a residential campus. Our empirical evaluation demonstrates the effectiveness of the proposed heuristics. An abridged preliminary version of the work appeared in [1].

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 describes our methodology including data collection and contextual separation of the data set. Section 4 presents two abnormal energy consumption detection heuristics. Section 5 discusses observations and results for each of the heuristics. Section 6 concludes the paper.

2 Related Work

With rapid development of smart grids, researchers have proposed various techniques for abnormal detection in smart grids. A sliding window framework was proposed to integrate historical sensor data and contextual features for detecting anomalous behavior [2]. Saad and Sisworahardjo [7] presented contextual anomaly detection algorithm to detect irregular power consumption using unsupervised learning algorithm and temporal context generated from meter readings. Rossi et al. [6] evaluated an approach for anomaly detection in smart grids derived from data streamed from smart meters. The proposed approach took into account the aspects of collective and contextual anomalies. In [5], the authors introduced load profiling methodology relying on consumption data from smart meters. The non-monitored users' load is then decomposed using artificial neural network trained with the available data.

Chen and Cook [4] transformed time series energy consumption data into a symbol sequence. They used suffix tree to identify occurrence of patterns. Then, the power sequences were clustered to find outliers in the data. They used two month data from apartments to detect energy outliers. We used clustering to get outliers in the data in one of the heuristics. Apart from clustering, we followed a completely different heuristic-based approach to detect abnormal behavior. In [8], the authors used peak energy consumption in daily readings to find abnormal energy consumption in buildings. They used energy consumption history and calculated variation from normal. They used mean and standard deviation to detect abnormal energy consumption. In our approach, we have also used energy consumption history. Apart from that we have used percentage change in consumption and distance from K-nearest neighbor days in heuristic 2 to detect abnormal behavior.

Bellala et al. analyzed data from 39 m in three buildings in a commercial campus of a large company [3]. They combined Support Vector Machines (SVM) with Hidden Markov Model (HMM) to propose a semi-supervised approach. They used occupancy models to reduce load on lighting on one floor. Our work is different as we used unlabeled data and unsupervised approach to detect abnormal consumption behavior.

3 Methodology

IIIT-Delhi campus is spread over 25 acres in Okhla, New Delhi, India. After phase 1 of construction, 30,000 m^2 of space was covered by buildings. Schneider Electric EM64XX series meters are deployed across campus and collect over 5 million data points every day. Over 200 of these smart meters are deployed inside 13 buildings in IIIT-Delhi campus. This includes student hostels, lecture halls, academic offices, faculty housing, library building (including student labs), sport arena, and centralized facilities. For our analysis in this paper, energy consumption data from 50 m of girls hostel and boys hostel were used.

3.1 Data Collection

Raspberry Pi controllers are connected to groups of meters over a common RS485 serial communication bus. Using smart controllers, the data is pulled from meters and transported over campus LAN. We utilized the Simple Measurement and Actuation Profile (sMAP) for data storage, transmission, and communication with devices.

Every 30 s, over 10 electrical parameters (including energy, voltage, power factor) are collected by each smart meter. This is in contrast to others, where data is collected at much larger granularity (30 min or above). Data collected at such high resolution allows detailed energy accounting. The data is stored into archive which is queried to access desired data. The dataset used contains energy data from January, 2014 to April, 2015 (16 months). All the days were labelled as weekday#, weekend#, holiday# in chronological order and meters were labelled as power# and backup# where # is the unique identifier (number).

Although, granularity as low as 30 s is useful for finding appliance usage signatures, very low granularity gives unnecessary level of details for energy consumption profiles. Therefore, we initially used aggregated data with hourly granularity to find abnormal energy consumption. Later, detailed causes for abnormal consumption were found using fine grained data.

3.2 Contextual Separation of Data

Energy consumption is dependent on the context. Separating and grouping data with similar context is helpful for enhanced data analysis. Therefore, we grouped the collected data on three parameters: hour of the day, type of the day, and type of supply.

In our case, the energy data is collected from student hostels of a residential campus. For regular students, the campus is functional eight months in a year. Rest of the four months includes winter and summer break of one and three months respectively. During breaks, many hostel wings are not occupied. Therefore, vacation and working days have different consumption patterns.

In IIIT-Delhi most of the classes are scheduled on weekdays. Weekdays and weekends have different consumption patterns as well. This is because hostel residents whose parents live nearby visit their homes during weekends. Thus, working days were further divided into weekdays and weekends. The three types of days are summarized as follows: Working days (Working weekdays and working weekends of the semester), Vacation days (Between semester holidays and mid-semester break), and Weekends (non-working).

To summarize, the data is divided into groups based on the following different contexts:

1. Day-wise context (3): Working days, weekends, and long holidays.
2. Hour of the day context (3): 0000–0600, 0700–1600 and 1700–2300 h.
3. Supply type context (2): Power and light backup.

In total, we formed 18 different groups ($3 \times 3 \times 2$ for each combination of day, hour and supply type). The expected intra-group daily energy consumption pattern is same. Therefore, abnormal energy consumption was analyzed by comparing different meters within the group.

4 Abnormal Consumption Detection Heuristics

We describe our heuristics to detect abnormal or unexpected energy consumption for a group of meters having same context. Using these heuristics, campus administrators can focus only on the anomalous meters and inspect unexpected behavior.

4.1 Heuristic 1: Abnormal Consumption Detection Using Percentage Change in Consumption

For a group of meters having same context, the energy consumption is similar. The actual energy consumption values depend on the behavior and usage pattern of the occupants. In IIIT-Delhi, most of the hostel rooms are occupied by same occupants for two consecutive semesters. Therefore, for a given meter, the change in behavior is dependent on external conditions including weather.

When we compare different meters having same external conditions, the rate of change in consumption is expected to be similar. The data used in this paper is collected from meters deployed at wing level inside two buildings. Every floor of the building has approximately three to five meters. Also, these buildings are located next to each other. In this environment, all meters are expected to exhibit similar behavior.

In other words, the rate of increase and decrease in energy consumption has to be similar for a given day. If most of the meters show decrease in their energy consumption, steep increase in consumption for the rest of the meters is an abnormal behavior.

This heuristic uses the percentage change of energy consumption to calculate abnormal energy consumption score. The input to the heuristic is a quantized set of energy consumption values. For a given day, the input to the heuristic is consumption data of current day and previous day of all meters. This data is chosen from one of the formed groups having same context. Using this heuristic, abnormal consumption score is calculated using the multiple steps. This data set is divided into subsets of a configurable width. For example, hours 0700–1600 can be divided into continuous subsets of length 5 namely 0700–1100, 0800–1200, 0900–1300, 1000–1400, 1100–1500, 1200–1600. Anomalous data for shorter durations of time is helpful to detect smaller segments of anomalies. This is because, while calculating percentage change in energy consumption, longer duration tends to normalize anomalous behavior. In case the abnormal behavior exists for longer durations, the effective score for longer durations will be magnified while combining the scores for a given set. This is because combined scores of subsets are used to find the score of a set. Also, shorter duration anomaly scores are helpful while narrowing down on abnormal behaviors.

For the given day and its previous day, total energy consumption is calculated. Using these values, the percentage change in energy consumption (P[i]) is calculated. To differentiate, this value is kept negative for decrease in total energy consumption and positive for increase in total energy consumption. From the values obtained in previous step, two sets of meters, namely increasing meters and decreasing meters, are formed. Meters with increase or decrease in energy consumption are called increasing meters or decreasing meters respectively.

Now, each of the meters in the two sets is multiplied with the ratio of meters in the opposite set to obtain a score S[i]. For example, if there are four meters in decreasing group and 1 m in increasing group. Then the percentage decrease for every meter in decreasing group will be multiplied with 0.2 (1/5) and the percentage increase for every meter in the increasing group will be multiplied by 0.8 (4/5). In this step, meters in minority will be multiplied with ratio of meters in majority and vice versa. The higher ratio of majority depicts higher abnormal behavior by the meters in the minority. The lower ratio of minority depicts lower abnormal behavior by the meters in the majority. Majority meters are those meters whose ratio among increasing and decreasing meters is higher. Similarly, minority meters are those meters whose ratio among increasing and decreasing meters is lower.

Finally, the score, S[i] is multiplied with average percentage change of the majority meters and the average percentage change of the minority meters. This is useful in comparing the anomaly score of current day with the anomaly score of other days. While comparing anomaly score of the current day all the values will be multiplied with the same factor. Therefore, for a given day this will not affect relative anomaly scores. Higher average of majority meters depicts higher chances of anomalous behavior for minority meters. Depending on the use case, one could also replace multiplication of scores with average percentage change of the majority and minority meters with their difference. As the sets were divided into subsets, the abnormal energy consumption score for a set could be calculated using the scores of subsets. To find the score for a given set, all its subset scores were averaged.

4.2 Heuristic 2: Abnormal Consumption Detection Using the Distance from K-Nearest Neighbor (KNN) Days

When all the energy consumption days are clustered together, outliers represent anomalous behavior. This is true when we assume the abnormal energy consuming days are significantly less in number compared to normal energy consuming days. In that case, the distance between two normal energy consuming days will be significantly lower than the distance between an abnormal energy consuming day and a normal day. Therefore, if we find the distance from K-nearest neighbouring days, the distance of anomalous days will be significantly higher than normal days. If there is a possibility of more than K-outliers to be close to each other, the distance from the mean of the data could be factored to calculate anomaly scores.

The input to the heuristic is consumption data. This data is chosen from one of the groups having similar context as the meter whose data is under consideration. A subset of data from a group is used to determine K-nearest neighbour days. Following are the configurations that we used to form subsets and find anomaly scores:

1. Consumption by all meters on all days: This is the basic configuration that can be used to find anomalous behavior. One should note that we are using the subset of data from a group out of the eighteen groups we formed. Therefore, by all meters on all days we mean the entire group. This method works well as data used have same context for the consumption data in a group. Also, one can further improve the results by selecting a subset of the data in a group to further narrow down the context. These cases are discussed in the remaining configurations.
2. Consumption by current meter on all days: This configuration can be used when some of the meters have very different consumption from the set. Different behavior from the set is not expected and should be investigated. If one observes such behavior, one could use this configuration to effectively find abnormal behavior. For example, if we consider the lifts in the hostels, these cannot be grouped with regular meters as the use case for the energy consumption in lifts is different than consumption in hostel rooms. Therefore, we can use this configuration to detect abnormal behavior in meters with unexpected or irregular consumption.
3. Consumption by all meters on current day: This is similar to consumption by current meter on all days. Instead of meters, the days having very different consumption from the rest of the days falls in this category. One can find anomalous energy consumption on some special days when the expected consumption is not same as other days in the group using this configuration. This also takes care of finding abnormal behavior when the data of only one day is given. This configuration works on the principle that per occupant consumption of anomalous meter will be farthest from rest of the meters for a given day.
4. Consumption by K-nearest neighbouring meters on all days: This is a special case in which only K-nearest neighbouring meters are used to find K-nearest

neighbour days. The meters in a group have same context for the consumption data. We can further narrow down the context by selecting K-nearest neighbouring meters for every meter. This can ensure that the neighbours used to detect anomalies are closer than other members in the group. This, in turn, ensures improvement in abnormal consumption scores.

5. Consumption by K-nearest neighbouring meters on current day: This case narrows down the context further for the previous case. This case is combination of case 3 and 4. One may choose this when conditions mentioned in both the cases 3 and 4 is true i.e. one needs a narrowed context for a special day.

One can directly use the sum of distances as anomaly scores. Also, one can normalize the scores to compare anomaly scores for different days or heuristics.

Table 1. Hourly consumption data (in Watt hour) for the groups.

Group#	Supply type	Day	Hours	Average usage	Max usage	Min usage
1	Backup	Holiday	Night hours	29.2754	322.6667	0.0035
2	Power	Holiday	Night hours	30.998	338.8333	0.0052
3	Backup	Holiday	Morning hours	29.9433	312.25	0.0035
4	Power	Holiday	Morning hours	29.4218	306.3333	0.0104
5	Backup	Holiday	Day hours	20.2426	317.625	0.0052
6	Power	Holiday	Day hours	24.3739	375.1667	0.0104
7	Backup	Weekday	Night hours	40.9852	415.4167	0.0026
8	Power	Weekday	Night hours	44.264	634.6667	0.0104
9	Backup	Weekday	Morning hours	41.8783	405.7917	0.0035
10	Power	Weekday	Morning hours	40.3534	312.2396	0.0052
11	Backup	Weekday	Day hours	28.843	357.9167	0.0026
12	Power	Weekday	Day hours	32.5202	350.5	0.0026
13	Backup	Weekend	Night hours	37.1992	387.75	0.0026
14	Power	Weekend	Night hours	39.2877	362.8333	0.0026
15	Backup	Weekend	Morning hours	38.808	369.1667	0.0069
16	Power	Weekend	Morning hours	37.4918	301.1667	0.0313
17	Backup	Weekend	Day hours	27.5047	335.9167	0.0026
18	Power	Weekend	Day hours	30.3037	358.8333	0.0052

5 Empirical Evaluation

To detect anomalies, energy data from January, 2014 to April, 2015 (16 months) is used for the analysis process. For the analysis, the energy consumption data from 50 m of girls hostel and boys hostel was used. For boys hostel, total covered area (on ground) and covered area (on floors) is 1116.19 m^2 and 6798.57 m^2 respectively. For girls hostel, total covered area (on ground) and covered area (on floors) is 838.99 m^2 and 3562.28 m^2 respectively. Girls and boys hostel have five and seven floors including ground floor. The hostel data is sampled with granularity of 30 s.

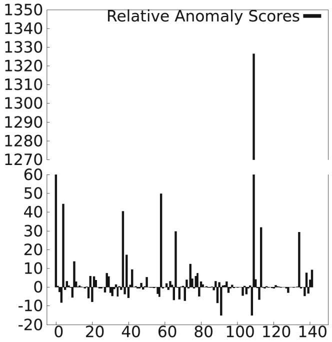

Fig. 1. Anomaly scores for heuristic 1 (light backup)

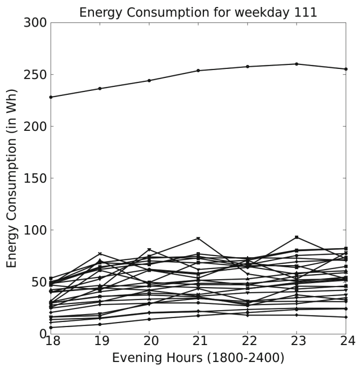

Fig. 2. High consumption in girls hostel ground floor wing BC

There are two power lines in the hostel namely power and light backup. Every floor has three wings namely A, B, and C. One energy meter collects the energy consumption data of 1/1.5/2 wings of a floor. For example, these could be (i) wing A or (ii) wing A with half of B or (iii) wing A and wing B. The hostels were fully occupied during semesters whereas only some of the floors are occupied during vacations. Therefore, the meters were contextually separated as discussed in Sect. 3.2. The expected behavior of meters within a group was same. This is because the meters are placed inside one building and therefore, the data is collected from exactly the same environment. Also, one should note that the factors like temperature, rainfall, humidity etc. do not play any role as we are comparing relative change in energy consumption for floors in same environment.

5.1 Observations and Results

We discuss observations and results describing effective usage for the heuristics described in the paper. The details about contextually separated groups is described in Table 1. The average and maximum hourly consumption for these groups varies between 20 to 45 and 301 to 635 W respectively. As one can notice that the energy consumption for some meters is twice when compared to others. Therefore, the formation of different contextual groups will help find anomalies effectively.

We next discuss the abnormal behavior highlighted by the heuristics mentioned in the paper.

Heuristic 1: Detection Using Percentage Change
The energy consumption data was directly fed into the heuristic to calculate anomaly scores. Figure 1 presents the anomaly scores for light backup. We can clearly identify several anomalous energy consumption trends (for example the weekday 111) using the heuristic. Upon further investigation, we found out that one of the meters located in girls hostel (ground floor wing BC) was displaying anomalous behavior as shown in Fig. 2. Consumption per occupant for this meter was found to be much higher compared to rest of the meters. This meter provides electricity to 6 occupants, 2 guest rooms, and a common room. Guest rooms are mostly unoccupied. The issue was discussed with the administrators and one of the three power supply phases was found using 4.9 A current (five times greater than the expected behavior which was shown by other two phases).

Such observations can be made real-time by campus administrators and misuse of energy can be pinpointed. Potential misuse of energy consumption includes use of personal appliances such as heater and refrigerators that are not permitted inside the hostel. This approach will also help in identifying sudden change in electricity consumption.

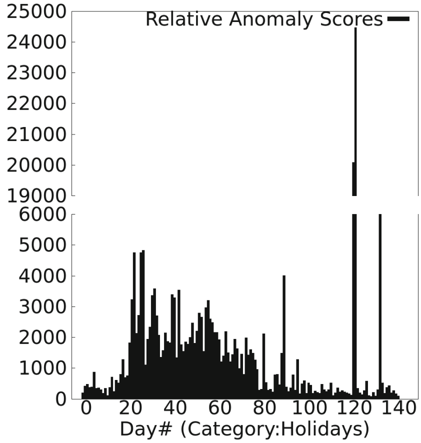

Fig. 3. Morning hours on holidays

Heuristic 2: Detection Using K-Nearest Neighbour Days
For heuristic 2, K = 10 was used to calculate the score. The anomalous days were far away from the given data set and thus, these were detected anomalous

by most of the configurations. The campus administrators can select multiple configurations and view the results to find abnormal consumption. Although most of the anomalies are listed by all the configurations, the result is configuration dependent. Also, the order of the anomalies and their priority depends on the configuration used. Therefore, we suggest combining output of all applicable configurations for finding anomalies.

Figure 3 represents anomaly scores for heuristic 2 using configuration 3. Abnormal consumption on day 120 is shown in all three zones: morning, day and night (only morning hours are shown here). Using real time monitoring of morning data, abnormal consumption in day and evening zones can be avoided.

6 Conclusions

We presented two heuristics to identify abnormal energy consumption using contextual grouping of smart meters. Groups were formed by collecting meters with same context. Similar expected behavior were used to compare energy consumption within the neighbourhood of the meters. The heuristics use percentage change in consumption and distance from K-nearest neighbour days to detect abnormal behavior. Data collected from fifty smart meters of hostels of IIIT-Delhi campus were used to analyze these heuristics. Multiple abnormal behaviors were found in the dataset. Our results show that the proposed heuristics successfully found abnormal energy consumption behavior. The abnormal behaviors were verified with facilities administrator of the campus. Our heuristics could be integrated into real-time energy monitoring systems to detect abnormal energy consumption.

References

1. Sial, A., Jain, A., Singh, A., Mahanti, A.: Profiling energy consumption in a residential campus. In: Proceedings of the CoNEXT Student Workshop, pp. 15–17 (2014)
2. Araya, D., Grolinger, K., ElYamany, H., Capretz, M., Bitsuamlak, G.: Collective contextual anomaly detection framework for smart buildings. In: Proceedings of the International Joint Conference on Neural Networks (2016)
3. Bellala, G., Marwah, M., Arlitt, M., Lyon, G., Bash, C.: Following the electrons: methods for power management in commercial buildings. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 994–1002 (2012)
4. Chen, C., Cook, D.: Energy outlier detection in smart environments. In: Proceedings of the 7th AAAI Conference on Artificial Intelligence and Smarter Living: The Conquest of Complexity (2011)
5. Ponocko, J., Milanovic, J.: Application of data analytics for advanced demand profiling of residential load using smart meter data. In: Proceedings of the 12th IEEE PowerTech Conference (2017)
6. Rossi, B., Chren, S., Buhnova, B., Pitner, T.: Anomaly detection in smart grid data: an experience report. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (2016)

7. Saad, A., Sisworahardjo, N.: Data analytics-based anomaly detection in smart distribution network. In: Proceedings of the International Conference on High Voltage Engineering and Power System (2017)
8. Seem, J.: Using intelligent data analysis to detect abnormal energy consumption in buildings. Energy Build. **39**(1), 52–58 (2007)