Georgia Journal of Science

Volume 74 Article 1

2016

Ranking Volatility in Building Energy Consumption Using Ensemble Learning and Information Entropy

Kunal Sharma
Georgia Institute of Technology, kunalsharma.usa@gmail.com

Jung-Ho Lewe
Georgia Institute of Technology, jungho.lewe@ae.gatech.edu

RANKING VOLATILITY IN BUILDING ENERGY CONSUMPTION USING ENSEMBLE LEARNING AND INFORMATION ENTROPY

Kunal Sharma[1], Jung-Ho Lewe[2]
Aerospace Systems Design Laboratory
Georgia Institute of Technology

I. Abstract

Given the rise in building energy consumption and demand worldwide, energy inefficiency detection has become extremely important. A significant portion of the energy used in commercial buildings is wasted as a result of poor maintenance, degradation or improperly controlled equipment. Most facilities employ sensors to track energy consumption across multiple buildings. Smart fault detection and diagnostic systems use various anomaly detection techniques to discover point anomalies in consumption. While these systems work reasonably well in detecting equipment anomalies over short-term intervals, further exploration is needed in finding methods that consider long-term consumption to detect anomalous buildings. This paper presents a novel approach for a multi-building campus to rank and visualize the long-term volatility of building consumption. This allows for the optimal allocation of limited time and resources for the detection and resolution of energy waste. The proposed method first classifies daily consumption into 5 classes using an ensemble learner and then calculates the information entropy on the resulting classification set to determine volatility. The ensemble learner receives input from a K-Nearest Neighbor classifier, a Random Forest classifier and an Artificial Neural Network. In general, buildings are expected to keep the same energy profile over time, all else being equal. Buildings that frequently change energy profiles are ranked and flagged by the system for review, which would call for the next step to reduce waste and costs and to increase the sustainability of buildings. Data on energy consumption for 132 buildings is obtained from energy management at the Georgia Institute of Technology. Experimental results show the effectiveness of the proposed approach.

Keywords: campus building, energy consumption, pattern recognition, ensemble learner, volatility detection

II. Introduction

Currently, commercial and residential buildings account for roughly 60% of the world's electricity consumption (UNEP 2020). Buildings alone spend 72 percent of U.S. electrical energy (DOE 2008). The demand for energy will continue to rise as a result of population growth, building comfort level improvement and increased demand.

Today's buildings are, however, widely reported to utilize energy inefficiently (Ardehali et al. 2003). Between 15% to 30% of the energy used in commercial buildings is

[1] Undergraduate Research Assistant, School of Computer Science.

[2] Research Engineer II, Aerospace Systems Design Laboratory (ASDL), School of Aerospace Engineering.

wasted by buildings that are deteriorating and maintained poorly with improperly controlled equipment desite having sensor and fault detection systems (Schein 2006) – (Katipamula et al. 2005). Overall, typical buildings consume 20% more energy than necessary due to faulty construction, malfunctioning equipment, incorrectly configured control systems and inappropriate operating procedures (Song et al. 2003) – (Wu et al. 2011). Moreover, anomalous events alone can account for 2% to 11% of the total energy consumption for commercial buildings (Heo et al. 2012). Focus of current fault detection systems on short-term intervals ignores the energy efficiency gains that can be made by considering long-term consumption patterns to identify anomalous buildings.

A. Literature Review

One of the most commonly applied techniques in the detection of abnormal electrical consumption is anomaly detection. Most applications of anomaly detection are specific to the target problem (Chandola et al. 2009). Various methods have been explored for processing electricity consumption data. An online contextual anomaly detection method is presented by Catterson et al. (2010) to find anomalies in sensor data which include loading, temperature, and the network configuration of transformers. A multi-agent system is described by McArther to find anomalies in the condition monitoring of electrical plant behaviors (McArther et al. 2005). Moreover, Jakkula compared different anomaly detection algorithms that identify anomalies in household energy consumption (Jakkula et al. 2010). In terms of the specific methods to find anomalies, recent papers often implement statistical-based, deviation-based, density-based and distance-based approaches. Many outlier procedures are based on using extreme studentized deviate (ESD) algorithm (statistical theory) and often achieve notable results as demonstrated in Seem (2007) and Liu et al. (2010).

The vast majority of current anomaly detection techniques applied to electricity consumption data are focused on the identification of specific point anomalies in the dataset. While point anomalies in consumption can be detected through these methods reasonably well, there has been little exploration in determining which buildings are anomalies in terms of their consumption patterns over a long period. Moreover, it would be in the interest of energy management to have a method rank anomalous buildings by priority in order to respond appropriately.

B. Problem Scoping

To fulfill the gap in identifying buildings with anomolous long-term energy consumption patterns, we present a method that first classifies daily consumption behavior then calculates the information entropy on the resulting classification set. To explore the practical application of our proposed method, we choose one specific multi-building campus as our case study. The Georgia Institute of Technology is chosen since it records the (kWh) energy consumption of each of its 132 campus buildings every 15 minutes.

Our research in this paper significantly differs from the previous work as it focuses on finding anomalous buildings based on consumption patterns instead of finding anomalous points of consumption. To that end, pattern classification and information entropy are employed as intermediary steps to visualize and rank long-term building consumption volatility.

III. Materials & Methods

A training dataset is used to create a model that classifies a building's consumption over 24 hours into one of 5 archetypical consumption patterns. These 5 archetypical consumption patterns were identified under the guidance of energy faculty at Georgia Tech. The training dataset contains 2,904 manually labeled classifications of energy consumption for all 132 buildings over 22 non-weekend and non-holiday days in 2017. The second dataset used contains the energy usage data for all 132 buildings from January 1st to August 31st, 2018. The classification model classifies every building's energy usage for 161 days, which excludes weekends and holidays. These classifications are used to identify the volatility of each building.

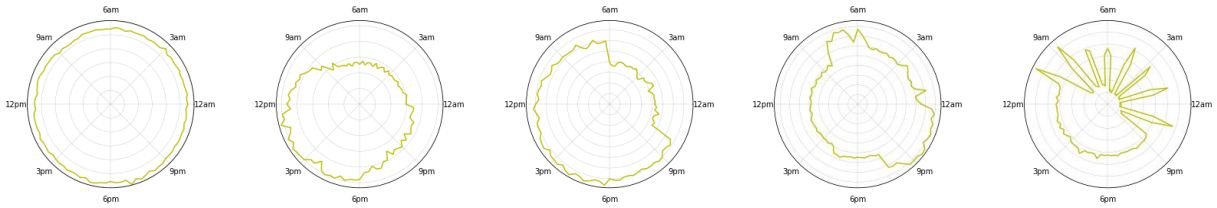
A. Variable Definitions

Energy management identified 5 distinct daily energy consumption patterns based on their expert knowledge and experience with the consumption behaviors of campus buildings. The energy consumption behaviors are labeled as concentric, people, scheduler, reverse and random.

Table 1: Definitions of Consumption Behavior Types

Behavior Type	Definition
Concentric	The building's energy consumption is relatively constant throughout the day.
People	The building uses more energy during work hours, likely due to energy consumption by people (lights, air-conditioning, heating, etc.).
Scheduler	The building that operated with a scheduler machine. Noted by sudden jumps and drops in energy usage at preset time periods.
Reverse	Energy usage during non-work hours is much higher than the energy usage for the rest of the day.
Random	The last catch-all term for erratic patterns that cannot be classified into the above four archetypes.

The random class helps classification models that otherwise would be forced to classify noisy data as the closest of the four well-defined archetypes. We can visualize the "energy signature" of these classification types by graphing a polar plot of energy consumption over a 24 hour period.



a. Concentric **b. People** **c. Scheduler** **d. Reverse** **e. Random**

Figure 1: Energy Consumption Types

B. Data Preprocessing

For both the training dataset and 2018 consumption dataset, only regular weekdays are considered; weekends and holidays are filtered out. Occasional sensor failure results in null values in the dataset. These values are replaced with the mean 15-minute consumption on that day for the building.

Both datasets are filtered to only include samples where the total energy consumption is higher than 125kWh in a day. A threshold is set to remove "Below Threshold" samples because preliminary experimentation finds that the energy consumption pattern of small buildings is erratic and hinders classification accuracy. Furthermore, after talking to faculty managers, we resolved to focus on larger buildings as they represent most of campus energy usage. After filtering samples with total consumption under 125kWh, the resulting training data has 1,076 instances and is unbalanced with "Concentric" representing 73.1% of the data. The filtered data is randomly split by an 80:20 train-to-test ratio resulting in 860 training instances and 216 test instances.

The training data is a matrix that has a column for every 15-minute interval in a day (96 columns) and a unique row for each building and day combination. Every cell contains a 15-minute average kW of the associated building on a particular day. The training data is normalized using min-max normalization so that the data is scaled to a value between 0 and 1. This accelerates the training and performance of the machine learning classifier.

The labeled (target) data is a matrix with each row containing the consumption classification for each corresponding row in the training data matrix. The labeled data matrix is one-hot encoded, meaning that it has 5 columns, one for each class. Only the column that corresponds to the classification will have a value of 1 while the rest have a value of 0. For example, a classification of 'Type 3' is represented as [0,0,1,0,0].

Data augmentation is a technique that increases the amount of training data by making duplicates with slight augmentations. This technique would improve our model's ability to generalize the 860 training instances to real-world data and reduce the chance of our model overfitting. To do this, the daily signature was "rotated" both forward and backward by shifting the data by 15-minute intervals 5 times. This encourages the model to generalize its understanding of energy signatures within a range of 75 minutes. We did not want to increase the range beyond 75 minutes as that could risk creating data that may not align with the real-world data. This data manipulation method created 10 additional instances of every training instance resulting in 9,460 training instances (860 original samples x 11 instances).

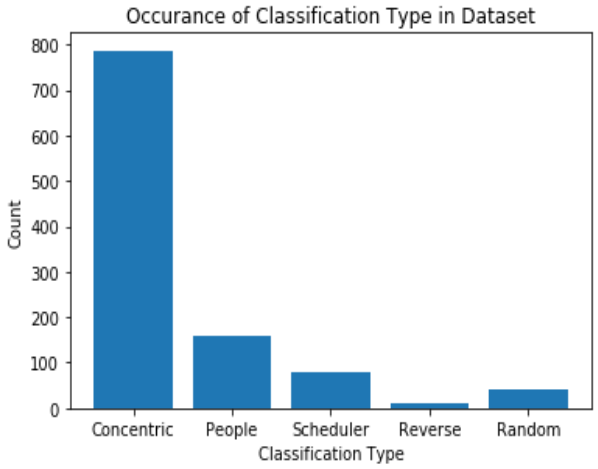


Figure 6: Class Frequencies Before Data Rotation

Table 2: Class Frequencies Before Data Rotation

Class	Count	Percentage
Concentric	787	73.1%
People	158	14.7%
Scheduler	80	7.4%
Reverse	11	1%
Random	40	3.7%
Total	1076	100%

C. Classification Algorithms

Classifying a building's daily energy consumption is an intermediate step to identifying its volatility. In preliminary classification experiments, the following classification models were tested: Decision Trees, K-Nearest Neighbors, Artificial Neural Networks, Naive Bayes, Logistic Regression, Random Forest, Support Vector Machine and Linear Regression. The highest performing models (Random Forest, K-Nearest Neighbors, Artificial Neural Networks) were combined to form an ensemble model. Sci-Kit Learn (sklearn) is a python-based statistics library that was used for the implementation of each of the models.

C.1 Random Forest Classifier (RF)

The Random Forest algorithm uses a collection of decision trees that "vote" on the classification of a data instance (Breiman 2001). A decision tree is a model that comes to a classification decision at the leaf node after splitting the data into multiple attributes. Data is split based on the attribute that best separates the observations by their target classes. In a random forest, several decision trees are formed from random subsets of the training data. The decision trees then "vote" on the classification of new data. Random Forests are preferred over a single decision tree as they are less prone to overfitting. Our classifier uses the default sklearn parameters except for 'n_estimators,' which is set to 200.

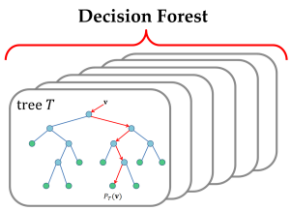


Figure 7: RF Classifier [15]

C.2 K-Nearest Neighbors (KNN)

The K-Nearest Neighbor algorithm looks at the K number of data points closest to the new data point (Guo 2012). The closest K data points to the new data point are identified based on their distance. The majority class of the K nearest points is the classification of the new data point. This method is commonly chosen for supervised classification tasks for its simplicity and intuitive approach. Our classifier uses the default sklearn parameters except 'k' is set to 5, 'p' is set to 2 and 'metric' is set to 'minkowski'.

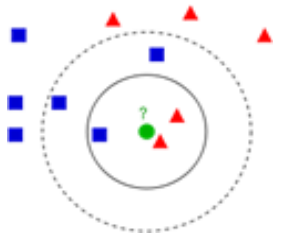


Figure 8: KNN Classifier [25]

C.3 Artificial Neural Networks (ANN)

An Artificial Neural Network is a network of nodes that propagate an input through a series of functions and output a classification (Sharma 2012). ANNs have multiple fully-connected layers. Data is propagated through the first layer, the input layer, then through a number of intermediary layers (hidden layers) before finally reaching the output layer for classification. After being given several training data instances, the weights across the network are adjusted through a process called backpropagation where the error is calculated at each layer. Typically more complex data with a larger input size will have more layers and more neurons. Our classifier has 6 layers with 96, 30, 20, 10, 8 and 5 neurons in each layer, respectively.

Figure 9: ANN Classifier [2]

C.4 Ensemble Model

Ensemble modeling is a process that uses multiple models to generate a final prediction on unseen data (Opitz 1999). Ensemble models are often used to bring down the prediction error of independent base models. Biases of any particular model are mitigated when multiple models work together. A voting classifier is a simple implementation of an ensemble model. Ensemble classifiers have been used in the medical field to improve breast cancer detection accuracy (Dubey 2019) and in the cybersecurity field to improve malware detection (Lu 2010). Our ensemble model uses the default sklearn parameters for a Voting Classifier model.

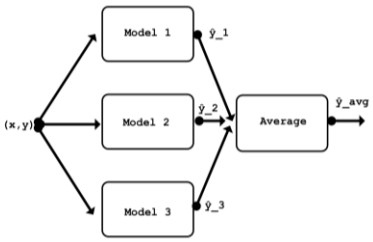


Figure 10: Ensemble Classifier [11]

D. Information Entropy Calculation

Information entropy is a measure of how much "information" there is in a set of data (Gary 2013). If all of the elements in a set of data are of the same class, the information entropy would be zero. If the set of data has more of a "mixture" of other classes, then the information entropy increases. The equation for information entropy is

presented below. The variable n represents the number of classes in the set of data. P_i represents the probability of the ith class occurring in the set of data.

$$H = \sum_{i=1}^{n} P_i \cdot \log_2 P_i \qquad \textbf{Eq. 1}$$

where the variable n represents the number of classes in the set of data, and P_i represents the probability of class i occurring in the set of data. A building's "volatility" score is determined by calculating the information entropy of its set of 24-hour energy consumption classifications over a long interval. Our definition of volatility is considering the entropy of a building's consumption *between* and not within 24-hour intervals since we are looking for buildings that are rapidly changing their consumption behavior over long intervals. Such buildings that are constantly switching between consumption archetypes will be scored as more volatile and are far more suspicious than buildings classified with a consistent consumption archetype.

IV. Results

A. Classification Accuracy

Each classifier is trained on the training data and evaluated on the test data. As expected, the ensemble model performs the best and will be the selected model.

Table 3: Classification Accuracies

Model	Accuracy
Random Forest	86.71%
K-Nearest Neighbor	85.38%
Artificial Neural Network	85.38%
Ensemble Model	**89.04%**

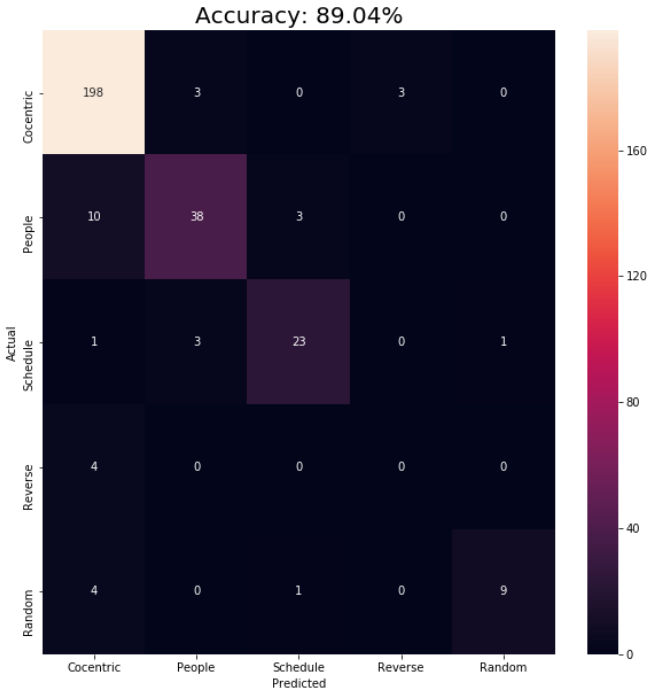


Figure 11: Ensemble Model Confusion Matrix

B. Classifying Building Consumption Behavior

Using the ensemble classifier, we can classify the 24-hour consumption for all 132 buildings from January 1st, 2018 to August 31st, 2018. Each classification is given a label (Table 4) for plotting.

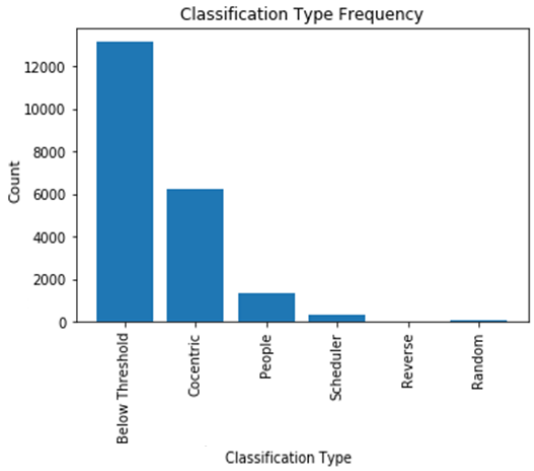


Figure 12: Class Type Frequency

Table 4: Class Type Frequencies

Label	Class	Count	Percent
-1	Below Threshold	13,157	62.1%
0	Concentric	6, 259	29.6 %
1	People	1,333	6.3%
2	Scheduler	341	1.6%
3	Reverse	5	0%
4	Random	77	0.3%
	Total	21,172	100%

While thresholding does considerably reduce the size of our dataset, the 24-hour consumption data points removed represent small buildings with relatively little consumption. By removing these erratic below-threshold data, the model's performance improves on buildings that represent a much larger portion of a campus' total consumption and energy costs.

C. Ranking Building Volatility by Information Entropy

We can calculate the entropy of the consumption behavior classifications for each of the 132 buildings. If the buildings that are consistently classified as below-threshold, they are filtered out. There remain 59 buildings. Figure 13 features a graph of these 59 buildings sorted by entropy.

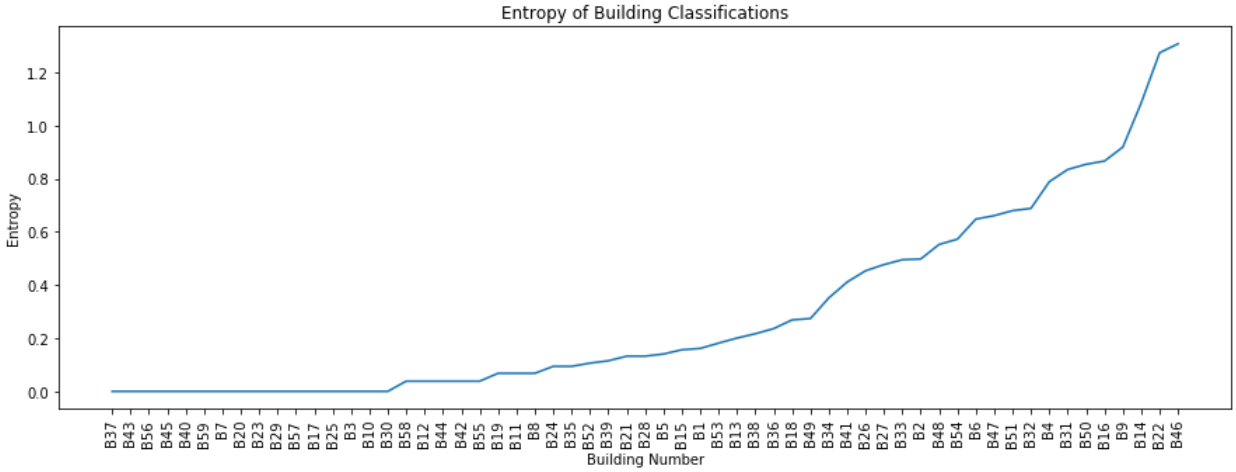


Figure 13: Sorted Entropy of Daily Energy Consumption Classifications for 132 Buildings

Buildings in Figure 13 are sorted by their entropy score. Building B37 has an entropy of 0, meaning that all its classifications are expected to be of a single type. Whereas building B46 has a high entropy of 1.3, meaning that its classifications should be heavily mixed. Graphs containing the classifications of B37 and B46 are plotted for the first 130 days (out of 161 total days) for visual clarity (Figure 14 & Figure 15). These graphs validate the respective entropy score for each building.

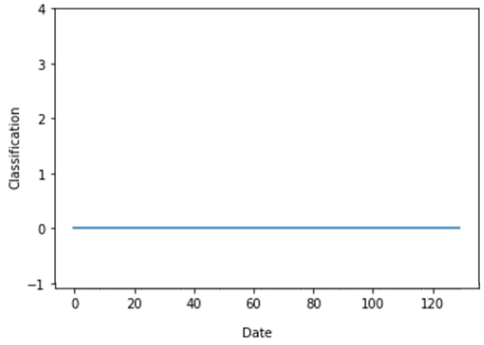


Figure 14: B37 Energy Consumption Classification

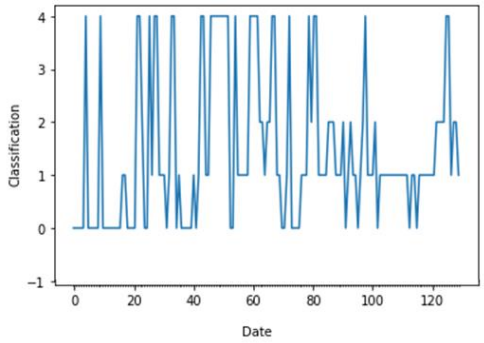


Figure 15: B46 Energy Consumption Classification

Using the entropy of the building's classifications, we now have a method to rank buildings by the highest volatility. A building with relatively high entropy such B46 can be flagged as it displays unusual energy consumption behavior. This gives energy management a means to prioritize investigations.

E. Visualizing Behavior Classification Entropy over Time

We can calculate the entropy of a building over a sliding window of time. This sliding window approach can summarize the behavior of a building over set time periods and help identify increases or decreases in entropy over time.

Below are the graphs of the entropy of buildings B37 and B46 calculated with a sliding window of four weeks. Building B37 still shows constant entropy over time. However, with building B46, we can see periods where entropy goes up and down. Energy management may be interested in understanding why such trends are present.

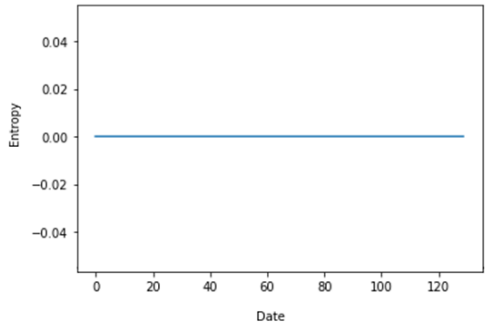


Figure 16: B37 28-day Sliding Window Entropy of Energy Consumption Classifications

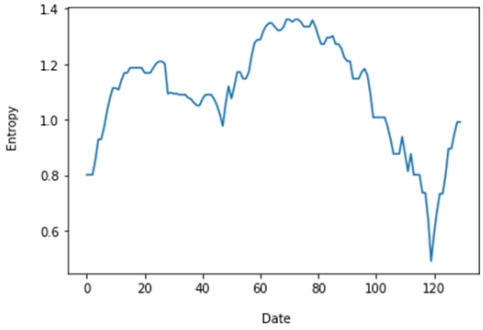


Figure 17: B46 28-day Sliding Window Entropy of Energy Consumption Classifications

We can visualize the changes in entropy with a sliding window for all buildings using a heatmap.

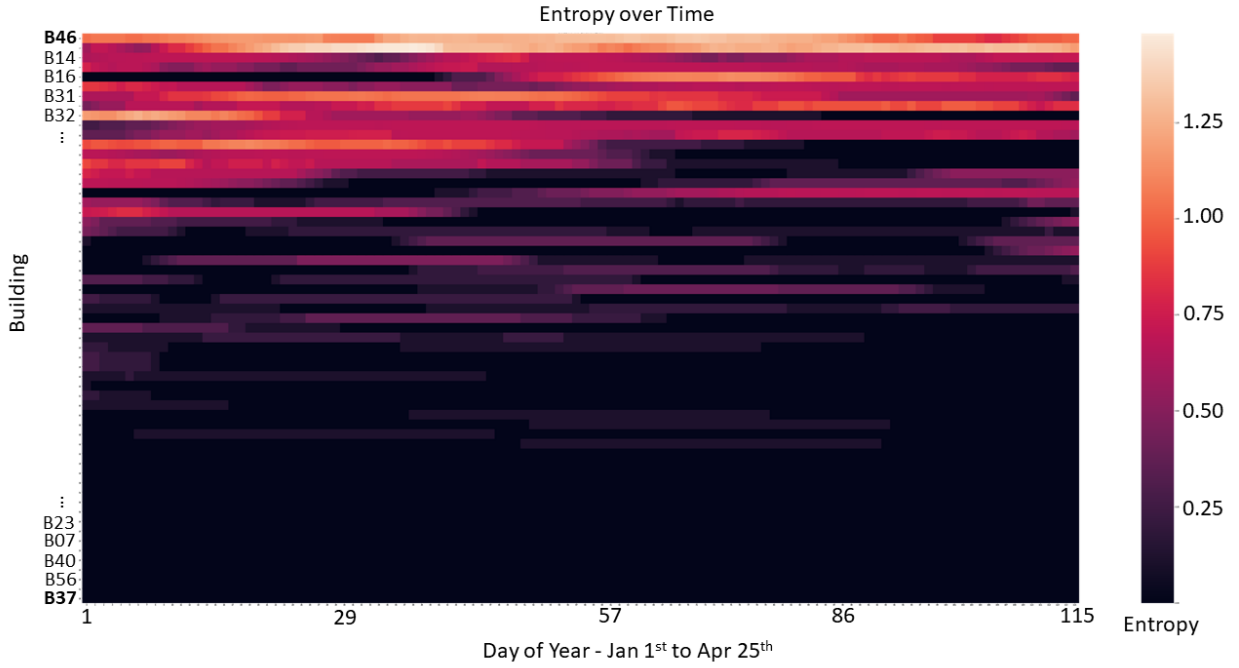


Figure 18: Entropy of Energy Consumption Classifications for all Above-Threshold Buildings calculated with a sliding window of 4 weeks.

V. Discussion

In conclusion, ensemble learning is an effective approach for classifying time series data. The ensemble learner included a random forest model, K-Nearest Neighbor model, and an artificial neural network model. The classification model and entropy calculations were used to summarize, rank and visualize the volatility of a building's energy consumption. These entropy calculations can be used to create a sliding window entropy heatmap of all buildings on a campus for energy management to identify and prioritize the investigation of buildings that have a higher risk of energy waste.

In the future, case studies can be conducted on the buildings that were identified by this algorithm to be of high risk. This feedback could be collected in a methodical way and integrated into the algorithm. Grid search can be used for parameter optimization for the ensemble model classifiers. More specific classifiers can be developed to classify the low-consumption cases that were filtered out. Further experiments can be conducted using sequential models for classification such as Recurrent Neural Networks which exhibits temporal dynamic behavior. Adding more examples for "Reverse" and "Random" to the data could help in balancing out the dataset.

VI. Acknowledgments

We would like to thank the Georgia Tech facility management for their domain expertise on the subject. We'd also like to thank Professor Dimitri Mavris, director of the Aerospace Systems Design Laboratory (ASDL), for providing this research oppurtunity.

VII. References

Ardehali M. M., and Smith T. F. 2003. Building Energy Use and Control Problems: An Assessment of Case Studies. ASHRAE Trans, 111–121.

Artificial Neural Network Diagram, www.upload.wikimedia.org/wikipedia/commons/4/46/Colored_neural_network.svg

Breiman L. 2001. "Random Forests," University of California, Berkeley, www.stat.berkeley.edu/~breiman/randomforest2001.pdf.

Catterson V.M., McArthur S.D., and Moss G. 2010. Online conditional anomaly detection in multivariate data for transformer monitoring. IEEE Trans. Power Deliv. 25, 2556–2564.

Chandola V., Banerjee A., and Kumar V. 2009. Anomaly detection: A survey, ACM Comput. Surv. (CSUR) 41, 15.

Dietterich T.G. 2000. Ensemble Methods in Machine Learning, Oregon State University, web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf.

Liu D., Chen Q., Mori K. and Kida Y. 2010. A Method for detecting abnormal electricity energy consumption in buildings. Journal of Computational Information Systems. 4887-4895.

Dubey K. 2019. Ensemble Classifier for Improve Diagnosis of the Breast Cancer Using Optical Coherence Tomography and Machine Learning. IOP Science. iopscience.iop.org/article/10.1088/1612-202X/aaf7ff.

Energy Efficiency for Buildings, Studio Collantin, United Nations Environment Program, www.studiocollantin.eu.

2008. "Energy Efficiency Trends in Residential and Commercial Buildings," Energy.gov, US Department of Energy. www1.eere.energy.gov/buildings/publications/pdfs/corporate/bt_stateindustry.pdf

Ensemble Learning Diagram. www.miro.medium.com/max/1400/1*fy-6esoTWsTutld4fdSyCQ.png

Gray R. 2013. Entropy and Information Theory. Information Systems Laboratory, Electrical Engineering Department, Stanford University. ee.stanford.edu/~gray/it.pdf.

Guo G. "KNN Model-Based Approach in Classification," School of Computing and Mathematics, University of Ulster, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.815&rep=rep1&type=pdf.

Heo Y., Choudhary R., and Augenbroe G. 2012. Calibration of building energy models for retrofit analysis under uncertainty, Energy Build. 47, 550–560.

2018. "Introduction to Random Forest: Blog: Dimensionless," Dimensionless Technologies pvt.ltd. dimensionless.in/introduction-to-random-forest/.

Jakkula V., and Cook D. 2010. Outlier detection in smart environment structured power datasets. In Proceedings of the 2010 Sixth International Conference on Intelligent Environments (IE), Kuala Lumpur, Malaysia, 29–33.

Katipamula S., and Brambley M. R. 2005. Methods for fault detection, diagnostics, and prognostics for building systems, A Review, Part I, HVAC&R Research 2005, 11 (1): 3 25.

Lu Y.-B. 2010. "Using Multi-Feature and Classifier Ensembles to Improve Malware Detection," Semanticscholar.org. pdfs.semanticscholar.org/d1fd/840c5a626e0004d2ca01847d6a97557c402a.pdf.

McArthur S.D., Booth C.D., McDonald, J., McFadyen I.T. 2005. An agent-based anomaly detection architecture for condition monitoring, IEEE Trans. Power Syst. 20, 1675–1682

Opitz, D. 1999. "Popular Ensemble Methods: An Empirical Study," Journal of Artificial Intelligence Research 11.

Schein J. 2006. "A Rule-Based Fault Detection Method for Air Handling Units." Energy and Buildings (Vol 38), 1485–1492.

Seem, J.E., 2007. Using intelligent data analysis to detect abnormal energy consumption in buildings. Energy and Buildings, 39 52-58.

Sharma, V. "A Comprehensive Study of Artificial Neural Networks," International Journal of Advanced Research in Computer Science and Software Engineering, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.9353&rep=rep1&type=pdf.

Song L., Liu M., Claridge D. E. and Haves P. 2003. Study of on-line simulation for whole building level energy consumption fault detection and optimization, Architectural Engineering 2003: Building Integration Solutions, 1-8.

Srivastava, T. 2019. "Introduction to KNN, K-Nearest Neighbors: Simplified," Analytics Vidhya. www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/.

Wu S., and Sun J. Q. 2011. Cross-level fault detection and diagnosis of building HVAC systems, Building and Environment, 46 (2011), 1558-1566.