Anomaly Detection Model of Time Segment Power Usage Behavior Using Unsupervised Learning

Wen-Jen Ho[1], Hsin-Yuan Hsieh[2], Chia-Wei Tsai[3]*

[1] Software Technology Institute, Institute for Information Industry, Taiwan
[2] Department of Computer Science and Information Engineering, National Taitung University, Taiwan
[3] Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taiwan
wenjen@iii.org.tw, 10911137@gm.nttu.edu.tw, cwtsai@nutc.edu.tw

Abstract

In Taiwan, the current electricity prices for residential users remain relatively low. This results in a diminished incentive for these users to invest in energy-saving improvements. Consequently, devising strategies to encourage residential users to adopt energy-saving measures becomes a vital research area. Grounded in behavioral science, this study introduces a feasible approach where an energy management system provides alerts and corresponding energy-saving recommendations to residential users upon detecting abnormal electricity consumption behavior. To pinpoint anomalous electricity usage within specific time segments, this research employs an unsupervised machine learning method, developing an anomaly detection model for the overall electricity consumption behavior of residential users. The model focuses on analyzing 2-hour intervals of electricity consumption, enabling more effective detection of abnormal usage patterns. It is trained using power consumption data collected from five actual residential users as part of an experimental study. The results indicate that the proposed anomaly detection model achieves performance metrics such as Precision, Recall, and F1-score of 0.90 or above, showcasing its potential for practical implementation.

Keywords: Energy saving, Electricity consumption behavior, Anomaly detection, Unsupervised learning

1 Introduction

Given the dwindling supplies of petrochemical energy and the escalating impacts of global warming, international governments and enterprises are increasingly focused on reducing dependence on petrochemicals, curbing carbon emissions, and conserving energy. Data from Taiwan's Bureau of Energy reveals that the industrial, service, and residential sectors are the three primary consumers of electricity, accounting for 55.9%, 17.7%, and 17.6% of Taiwan's total electricity consumption, respectively. Although the residential sector ranks third, its consumption is nearly on par with the service sector. Thus, effectively curbing electricity consumption in the residential sector could yield significant energy savings. However, a challenge arises as, unlike the predominantly medium- to large-scale users in the industrial and service sectors, the residential sector consists of numerous small-scale users, totaling about 14.23 million households. Each of these households exhibits distinct electricity consumption patterns, making it challenging to devise a one-size-fits-all energy-saving mechanism for the entire residential sector.

With the rapid development of information and communication technology and Internet of Things devices in recent years, smart meters and home energy management systems have become increasingly popular, leading to year-by-year reductions in the cost of collecting residential users' electricity consumption data. If electricity consumption data can be analyzed to further provide residential users with multiple energy-saving services, residential users will become more willing and motivated to implement energy conservation strategies. However, the electricity prices for residential users in Taiwan are currently very low, and implementation of energy-saving strategies has not been enhanced. Therefore, other mechanisms should be utilized in tandem to increase user-willingness in implementing energy-saving strategies. According to [1], the integration of behavioral science into energy-saving technology can yield good energy-saving benefits. Techniques related to behavioral science include: 1) goal-setting theory, 2) avoiding information asymmetry, 3) loss aversion, and 4) social comparison theory. Loss aversion means that people's aversion to losses is stronger than their liking for gains, so avoiding losses is often the priority when people take action. Therefore, energy-saving services can integrate the tracking of electricity consumption and detection of users' abnormal electricity consumption behavior and alert users, emphasizing on unnecessary losses caused by abnormal power consumption behavior. Moreover, users' loss aversion psychology should be leveraged to enhance their willingness to implement energy-saving services. To this end, identifying the abnormal electricity consumption behavior of residential users is the main challenge.

To overcome the challenge mentioned above, [2-5] have established a mechanism for detecting abnormal electricity consumption through a regression model. However, the

*Corresponding Author: Chia-Wei Tsai; E-mail: cwtsai@nutc.edu.tw
DOI: 10.53106/160792642024052503011

setup of the regression model requires complete physical eigenvalues (e.g., temperature and humidity) of users, increasing the setup cost. Some studies [6-7] have used supervised machine learning to develop anomaly detection models. However, the anomaly label is hard to obtain, and is often irrelevant to unseen anomalies. Therefore, the feasibility of supervised learning-based models for detecting abnormal electricity consumption behavior is low in actual application environments. [8] describes an unsupervised learning strategy implemented in a model for detecting abnormal electricity consumption behavior, but this model only analyzes the electricity consumption of commercial buildings. Refs. [9-10] also utilize machine learning strategies to propose anomaly detection mechanisms for the road lighting system and the industrial electrical system, respectively. Additionally, our previous study [11] proposes a model for detecting residential users' abnormal electricity consumption behavior by combining an unsupervised learning model with a high-level feature extraction technique. This model only focuses on the analysis of daily electricity abnormal consumption behavior, and thus the model proposed by [11] cannot provide energy-saving services in a timely manner. In view of this, this study aims to use unsupervised learning to construct a model for detecting abnormal electricity consumption behavior in specific time segments and analyze electricity consumption behavior every two hours to improve the detection timeliness of abnormal electricity consumption. To verify the feasibility of the proposed model in an actual application environment, we conduct modeling and experiments based on the electricity consumption data of five actual residential users, and analyze and evaluate the experimental results. The reminder of this paper is organized as follows: Section 2 reviews the relevant literature and techniques; Section 3 introduces the research method and process; Section 4 explains and discusses the implementation method and results; Section 5 summarizes the conclusions of this paper, as well as future research topics and directions.

2 Background

Detection of outliers or anomalies in data has been studied by the statistical community since the 19th century [12]. An anomaly is generally broadly defined as a situation in which the patterns in data do not conform to the intentions clearly defined in normal behavior [13]. Anomalies can also be referred to as outliers, deviations, inconsistencies, or exceptions. In addition, anomaly detection as a major category of novelty detection evaluates whether a new sample or instance uses a different pattern from the past dataset (the dataset used in the training model). Novelty detection has two categories. The major category of novelty detection is outlier detection, which identifies whether there are some outliers in the current dataset significantly different from other data. In addition, anomalies can be divided into three types by anomaly interpretation and identification methods. The following table summarizes the descriptions of the three different types of anomalies.

Table 1. Classification of anomalies

Type	Description
Point anomalies	If a single data instance is found to be different from the remaining data instances, then this instance is called a point anomaly.
Contextual anomalies	If a single data instance falls within the normal range but is abnormal in a specific context, it is called a contextual anomaly. Contextual attributes include space-time attributes and behavioral attributes.
Collective anomalies	If a data collection is anomalous compared to the entire dataset, it is called a collective anomaly. In other words, data instances alone are not anomalous, but they are anomalies when they aggregate together.

The occurrence of contextual anomalies depends on the availability of contextual attributes in data. Point anomalies or collective anomalies can also be contextual anomalies if they are analyzed in specific contexts. Specifically, based on contextual information, a point or collective anomaly detection problem can be converted into a contextual anomaly detection problem. Therefore, most studies transform anomaly events into contextual anomalies for analysis and processing.

The trained anomaly detection models can be divided into three types according to whether the training dataset has labeled data. The description of each type is as follows:

- Supervised anomaly detection model: Both normal and abnormal data are labeled in the training dataset, so a binary classifier can be used to detect outliers. However, this strategy is characterized by high cost of obtaining labels, and it is difficult to find accurate and representative labels. In addition, lower numbers of anomaly labels in practical scenarios often cause data imbalance, which makes this training strategy more challenging.
- Semi-supervised anomaly detection model: Only normal data are labeled in the training dataset, so this strategy is more widely used than the supervised strategy. This strategy uses a one-class classifier to construct an anomaly detection model, but the major issue is that when the model detects an anomaly, it is impossible to determine whether the detection is accurate or whether it is some unseen normal event during training.
- Unsupervised anomaly detection model: Both normal and abnormal data are unlabeled, so it is necessary to construct a model based on certain statistical assumptions, such as the assumption that the probability of anomalies is very low. However, if the assumption is not true, the model will have a high false positive rate. In addition, the identified anomalies have low interpretability, and need to be explained by experts of the related subject.

In addition to the above three trained anomaly detection models, [2] also classifies anomaly detection techniques into five types: 1) probabilistic, 2) distance-based, 3) reconstruction-based, 4) domain-based, and 5) information-theoretic. Probabilistic anomaly detection generally uses a statistical hypothesis test to diagnose whether the observed data is anomalous. Distance-based anomaly detection determines whether a novel event is an anomaly according to the distance from the normal event. When performing anomaly detection, the k-nearest neighbor algorithm, for example, first finds the average distance between all the data and their nearest k neighbors to set a threshold. If the average distance between the detected data and their neighboring points is greater than the threshold, the detected data will be regarded as anomalies; otherwise, they will be regarded as normal data. Reconstruction-based anomaly detection inputs a novel event into the trained machine learning model (e.g., regression or classification model), and judges whether it is an outlier based on the differences between the model output (reconstructed data) and actual value. Domain-based anomaly detection uses the training data to define a boundary around the normal data, and detects whether the training data is within the boundary. The training data beyond the boundary are regarded as anomalies. Information-theoretic anomaly detection analyzes the amount of information in the dataset based on various metrics such as entropy, relative entropy, and mutual information, and detects anomalies by analyzing the difference in the amount of information between the novel event and original event.

In the detection of abnormal electricity consumption behavior, Zhang et al. [3] have developed a reconstruction-based anomaly detection model. They have constructed a linear regression model for electricity load, and then used the prediction result as the baseline. If the actual electricity consumption data is significantly lower or higher than the baseline, the event is considered an anomaly. Zhou et al. [4] also proposed a reconstruction-based anomaly detection model, and improve the accuracy of user load forecasting by integrating two regression models, i.e., autoregressive integrated moving average and artificial neural network models, thereby increasing the accuracy of electricity consumption anomaly detection. Luo et al. [5] proposed an anomaly detection model based on dynamic regression results that calculates an active adaptive threshold instead of using a fixed threshold to detect anomalies in the difference between the predicted result and actual load.

In addition to reconstruction-based anomaly detection models, Jokar et al. [6] have developed a power theft anomaly detection model based on supervised learning. During the training process, this model uses the *k*-means clustering algorithm to extract primary electricity consumption behavior, and then adopts the support vector machine algorithm to train a binary classifier. Pinceti et al. [7] have conducted a comparative study on the model performance to analyze the performance of different supervised learning models in detecting electrical load anomalies. Anomaly detectors based on supervised classification models usually perform excellently in anomaly detection, and the detection is characterized by high rationality and interpretability. However, this method is disadvantaged by its high training cost and poor applicability, and often encounters two challenges: 1) the training of high-efficiency supervised learning requires high-quality datasets; however, collecting high-quality data takes a long time and is expensive. 2) Labeled datasets may not necessarily represent the future events, so the dataset needs to be updated in real-time. In view of this, Fan et al. [8] proposed an anomaly detection model of electricity consumption in buildings based on unsupervised classification to reduce the training cost of anomaly detection based on supervised learning. In this study, we employ spectral density analysis to obtain users' main load frequency, apply the decision tree to determine the main eigenvalues that affect the electricity consumption behavior, and finally calculate the anomaly score of each observation using an unsupervised learning model, i.e., autoencoder.

3 Research Methods

This section discusses the research steps and process. We use the data of five actual low-voltage residential users from August 2021 to June 2023 for analysis. Using the proposed model that detects residential users' electricity consumption anomalies based on the domain-based anomaly detection mechanism, we analyze the electricity load every two hours to identify whether users' electricity consumption behavior during this time segment is normal. The research methods and steps in this study are described as follows (as shown in Figure 1).

3.1 Data Preprocessing

Electricity consumption data of low-voltage users are collected mostly by smart meters and home energy management systems. However, the data transmission process may be affected by noise, or the sensor may malfunction temporarily, resulting in noise, outliers, or missing values in the collected data. Noise and anomaly data will affect subsequent model training and performance. Therefore, before data analysis and model training, it is necessary to filter and process problematic data samples. For local data containing missing values, we calculate the average value of other days in the same time segment, and use such values to fill in all missing values. If the number of missing values exceeds 10% of the dataset for that day, the electricity load data for that day is deleted. Since we use an unsupervised learning strategy, it is necessary to filter the outliers in the dataset to ensure no noise and abnormal electricity consumption behavior in the training dataset. To this end, we use the Isolation Forest model for outlier detection and removal. Specifically, the extreme values (approximately 5%) in the normal distribution are removed as outliers. Since the international electricity industry samples users' electric load at intervals of 15 minutes, we re-sample the original load data using the average method. In addition, the values are rounded to the second decimal place, and the unit is converted from watts (W) to kilowatts (KW). In addition to the eigenvalues of electricity consumption data, we also query data from Taiwan's Central Weather Bureau based on user addresses, and integrate hourly temperature and humidity data with users' electricity consumption data. Date information is also one of the key factors affecting electricity consumption behavior. Therefore, we analyze the date of the electricity consumption data to collect the characteristic date information (e.g., the month, day of the week, and whether it is a weekday) of this electricity consumption time segment, and then integrate these into the data.

3.2 Contextual Analysis and Data Augmentation

In Taiwan, temperature is the main factor affecting users' electricity consumption behavior. Therefore, we use a temperature of 26°C as the classification threshold. Months with a temperature above 26°C are classified as summer months, and months with a temperature below 26°C are classified as non-summer months. Modeling and training are conducted for different electricity consumption scenarios. We divide the dataset according to the 8/2 rule, that is, 80% of the dataset as the training set and 10% as the test set. A very small amount of data are available for model training and testing due to system disconnection. To evaluate the performance of the electricity consumption anomaly detection model, we multiply the electricity consumption load data by a small random value (0.97–1.03) to augment the dataset. In addition, we use the data of simulated abnormal electricity consumption behavior to evaluate whether the model can detect abnormal electricity consumption events. The simulated abnormal electricity consumption behavior includes:

- High abnormal electricity consumption: The user's highest electricity consumption load is multiplied by 1.8–2.5.
- Low abnormal electricity consumption: The user's highest electricity consumption load is divided by 1.8–2.5.
- Abnormal electricity consumption showing steep rise and steep drop: High abnormal electricity consumption load and low abnormal electricity consumption load are mixed to simulate the abnormal electricity consumption behavior of steep rise and steep drop.

The training dataset only contains the data of normal electricity consumption behavior. However, the test dataset contains the data of both normal and abnormal electricity consumption behavior, accounting for 70% and 30%, respectively.

3.3 Data Standardization and Feature Extraction

To eliminate the impact of the scale of each feature on the model training process and accelerate the learning speed of the model, we use standard scaling to standardize the data, convert the data distribution to a normal distribution, and eliminate feature noise to facilitate subsequent modeling and analysis. We introduce principal component analysis to reduce the data dimensionality while retaining most of the training data. Maximum likelihood estimation is applied to select a suitable dimensionality reduction, so as to achieve dimensionality compression while retaining sufficient information.

3.4 Model Training

An unsupervised local outlier factor (LOF) model is employed for anomaly detection. Based mainly on density, this model not only considers whether the data point is far from the center of the entire dataset, but also considers the local density around the data point. Even if the anomaly data is close to the data center, this algorithm can still effectively find the anomaly data with lower density in the surrounding region. Scikit-learn is used for modeling. During the training process, we use the preset "auto" to set the ratio of pollution values to ensure that the input data are all normal electricity consumption data. During the model fine-tuning process, we set the ratio to (0.01,0.05) to select the optimal parameters.

3.5 Performance Evaluation and Fine-tuning

We use three common metrics to comprehensively evaluate the performance of the trained model: Precision, Recall, and F1-score. If the model performance is found to be lower than expected, we return to the stage of model training and fine-tune the model. During model fine-tuning, we use the Optuna [14] suite to tune the hyperparameters of the model. Through various optimization strategies (including grid search, random search, and tree-structured Parzen estimator algorithm) provided by the Optuna suite, we estimate the optimal parameter combination for the model to ensure performance and generalization capabilities of the time-segment anomaly detection model. Finally, the effectiveness of the established model is verified using the test dataset.

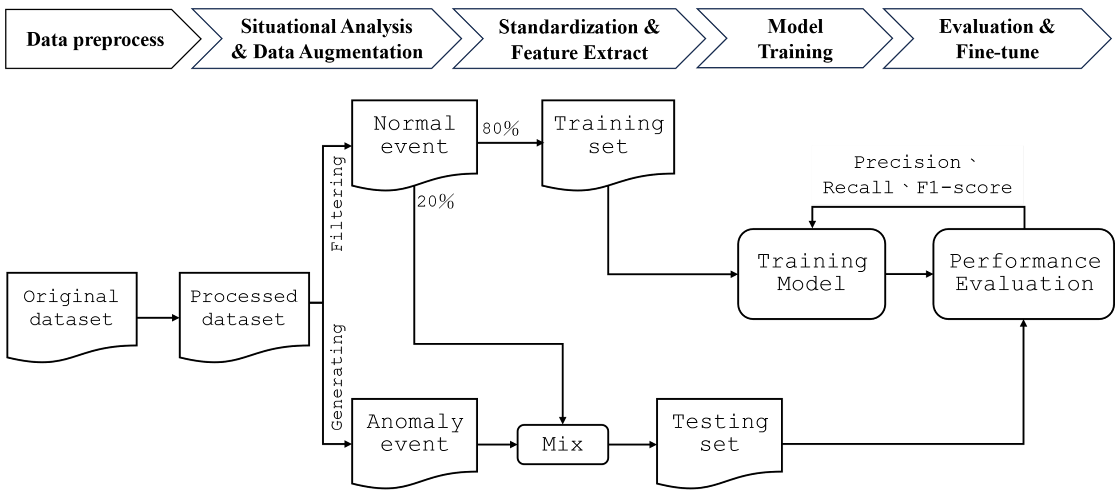

Figure 1. Flowchart of model for detecting abnormal electricity consumption behavior

4 Experiment Results and Discussion

This study uses data from August 2021 to June 2023 for analysis. Due to the outbreak of the COVID-19 pandemic during the analysis period, the electricity consumption behavior of residential users at some experimental sites was affected by remote work and home isolation, and the electricity consumption behavior significantly changed. To avoid the interference of all special electricity consumption behaviors caused by COVID-19 on future anomaly detection, we selected five users whose electricity consumption behavior was less affected to perform modeling and evaluation. The curves of the five users' electricity consumption loads are as shown in Figure 2 to Figure 6.

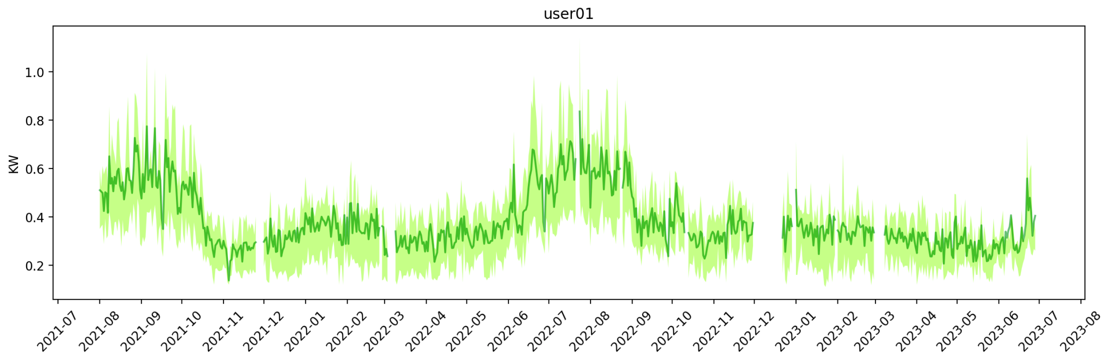

Figure 2. Electricity consumption of User01

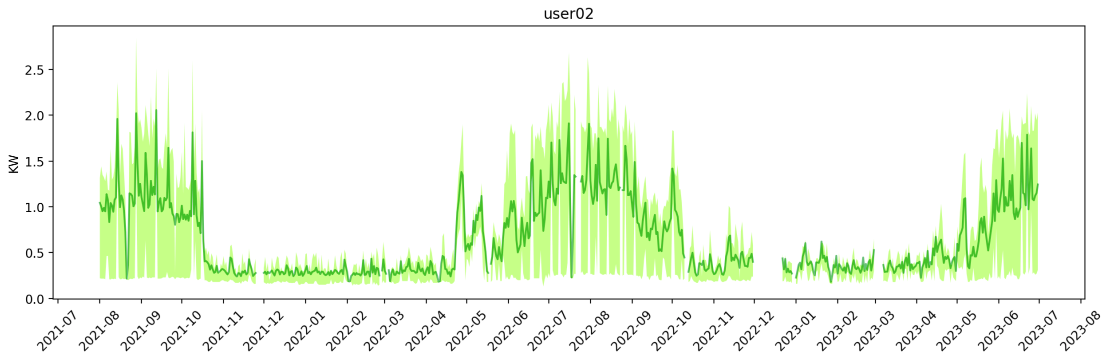

Figure 3. Electricity consumption of User02

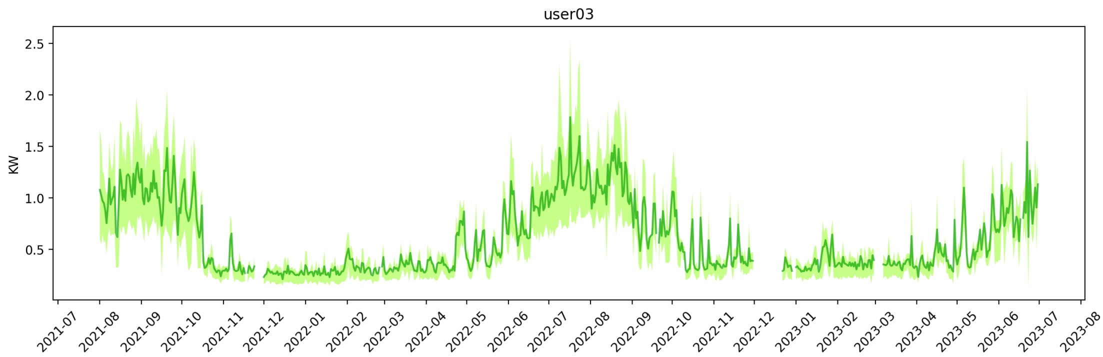

Figure 4. Electricity consumption of User03

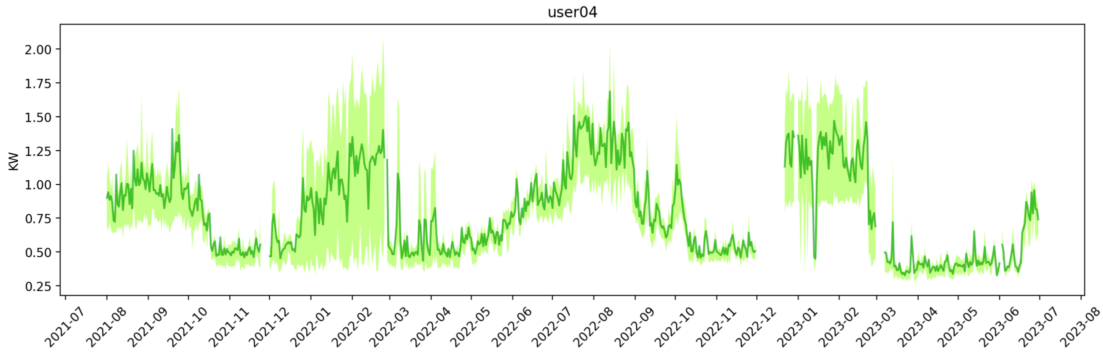

Figure 5. Electricity consumption of User04

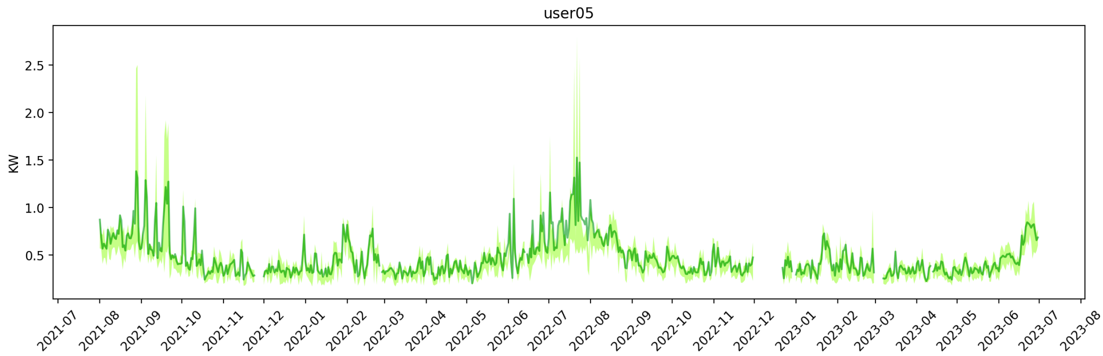

Figure 6. Electricity consumption of User05

From the electricity consumption curve of each user, we can observe missing values for all five users in November 2022, December 2022, and March 2023, which may be caused by equipment maintenance. July and August 2022 witnessed electricity consumption peaks, possibly due to the frequent use of air conditioners with high electricity consumption in summer. At the end of January and the beginning of February in 2023, the frequency of heating use increased due to lower temperatures, so electricity consumption peaks were also witnessed during this period. However, due to climate warming in recent years, the temperature did not drop significantly in winter, and it is difficult to see obvious differences in the electricity consumption curve.

After the contextual analysis, we considered 26°C as the baseline temperature. In the data, the months with temperatures higher than 26°C were designated as summer months, and those less than 26°C were specified as non-summer months. A model for detecting abnormal electricity consumption time segments was established using the LOF algorithm for both summer and non-summer months. We fine-tuned the poorly-performing models to improve their performance, and used Precision, Recall, and F1-score to evaluate the model performance. The performance of the anomaly detection models for each user in summer and non-summer months are summarized in Table 2 to Table 7 and Table 8 to Table 11, respectively. Here, Tx represents the time segment between x and x+2, e.g., T2 indicates the time segment 2:00 AM–4:00 AM. Through experiments on five residential users, we observed that the average F1-score of the model for each user all exceeded 0.9, showing the excellent performance of the model in identifying abnormal time segments. However, we also noticed that some Precision values were too high. This indicates that some real positive samples were mistaken for negative samples in the prediction of the model, resulting in false negatives. In some cases, this may lead to actual anomalies not being detected in a timely manner. This is especially true for the model for User04 in non-summer months, where the model was too cautious in judging abnormal electricity consumption. While avoiding anomaly misjudgment, this model fails to detect the real abnormal electricity consumption behavior, which will affect the model performance.

Table 2. Model performance in non-summer months for User01

	Precision	Recall	F1-score
T0	0.99	0.99	0.99
T2	1.00	0.87	0.93
T4	1.00	0.89	0.94
T6	0.94	0.93	0.94
T8	0.89	0.89	0.89
T10	1.00	0.91	0.96
T12	0.89	0.90	0.89
T14	0.93	0.91	0.92
T16	0.95	0.92	0.93
T18	0.89	0.92	0.90
T20	0.97	1.00	0.99
T22	1.00	0.91	0.95

Table 3. Model performance in summer months for User01

	Precision	Recall	F1-score
T0	1.00	0.89	0.94
T2	1.00	0.96	0.98
T4	0.84	0.94	0.88
T6	0.91	0.97	0.94
T8	0.94	0.94	0.94
T10	0.97	0.97	0.97
T12	0.92	0.92	0.92
T14	1.00	0.88	0.94
T16	1.00	0.96	0.98
T18	1.00	0.81	0.89
T20	1.00	0.89	0.94
T22	1.00	0.84	0.91

Table 4. Model performance in non-summer months for User02

	Precision	Recall	F1-score
T0	1.00	0.89	0.94
T2	1.00	0.90	0.95
T4	1.00	0.93	0.96
T6	0.93	0.90	0.91
T8	0.90	0.90	0.90
T10	0.91	0.89	0.90
T12	0.91	0.89	0.90
T14	1.00	0.93	0.96
T16	0.80	0.87	0.84
T18	0.98	0.91	0.95
T20	1.00	0.87	0.93
T22	1.00	0.88	0.94

Table 5. Model performance in summer months for User02

	Precision	Recall	F1-score
T0	1.00	0.79	0.88
T2	0.91	0.86	0.89
T4	0.82	0.97	0.89
T6	1.00	0.94	0.97
T8	1.00	0.91	0.95
T10	0.98	0.89	0.93
T12	0.94	0.92	0.93
T14	1.00	0.86	0.92
T16	0.91	0.90	0.91
T18	0.98	0.91	0.95
T20	1.00	0.87	0.93
T22	1.00	0.87	0.93

Table 6. Model performance in non-summer months for User03

	Precision	Recall	F1-score
T0	0.90	0.90	0.90
T2	1.00	0.93	0.96
T4	1.00	0.88	0.94
T6	0.92	0.88	0.90
T8	0.84	0.90	0.87
T10	1.00	0.90	0.95
T12	1.00	0.91	0.95
T14	0.95	0.94	0.95
T16	0.92	0.93	0.92
T18	1.00	0.90	0.95
T20	1.00	0.90	0.95
T22	0.93	0.91	0.92

Table 7. Model performance in summer months for User03

	Precision	Recall	F1-score
T0	1.00	1.00	1.00
T2	1.00	0.88	0.94
T4	1.00	0.99	0.99
T6	1.00	0.98	0.99
T8	1.00	0.91	0.95
T10	1.00	0.88	0.93
T12	0.87	0.86	0.87
T14	1.00	0.90	0.95
T16	1.00	0.88	0.94
T18	0.89	0.80	0.84
T20	1.00	0.89	0.94
T22	1.00	0.99	0.99

Table 8. Model performance in summer months for User04

	Precision	Recall	F1-score
T0	0.97	0.92	0.95
T2	0.88	0.91	0.90
T4	0.95	0.90	0.93
T6	0.94	0.87	0.90
T8	0.91	0.87	0.89
T10	0.83	0.83	0.83
T12	1.00	0.82	0.90
T14	0.88	0.94	0.91
T16	1.00	0.80	0.89
T18	0.92	0.94	0.93
T20	0.93	0.93	0.93
T22	1.00	0.94	0.97

Table 9. Model performance in summer months for User04

	Precision	Recall	F1-score
T0	1.00	0.95	0.97
T2	1.00	1.00	1.00
T4	1.00	1.00	1.00
T6	1.00	0.95	0.98
T8	1.00	0.94	0.97
T10	1.00	0.94	0.97
T12	1.00	0.89	0.94
T14	1.00	0.89	0.94
T16	1.00	0.92	0.96
T18	1.00	0.94	0.97
T20	1.00	0.90	0.95
T22	1.00	0.95	0.97

Table 10. Model performance in non-summer months for User05

	Precision	Recall	F1-score
T0	0.94	0.90	0.92
T2	0.91	1.00	0.95
T4	1.00	0.91	0.95
T6	0.87	0.90	0.89

T8	0.90	0.87	0.88
T10	0.95	0.87	0.91
T12	0.97	0.91	0.94
T14	0.97	0.90	0.93
T16	1.00	0.91	0.95
T18	0.89	0.90	0.90
T20	0.95	0.88	0.91
T22	1.00	0.90	0.95

Table 11. Model performance in summer months for User05

	Precision	Recall	F1-score
T0	1.00	0.97	0.98
T2	1.00	0.96	0.98
T4	1.00	1.00	1.00
T6	1.00	0.93	0.96
T8	0.92	1.00	0.96
T10	1.00	0.86	0.93
T12	1.00	0.89	0.94
T14	1.00	0.93	0.96
T16	1.00	0.90	0.95
T18	0.95	1.00	0.98
T20	1.00	0.92	0.96
T22	1.00	0.97	0.98

5 Conclusions

This study employs an unsupervised domain-based anomaly detection technique to construct a model that identifies abnormal electricity consumption time segments for residential users, and it assesses the viability of the proposed strategy. For the experiment, we selected five residential users who had comprehensive data and were unaffected by the COVID-19 pandemic for analysis, modeling, and training. Data was gathered from August 2021 to June 2023. During Taiwan's onset and management of the COVID-19 pandemic, the electricity consumption patterns of residential users shifted notably, characterized by extended home stays and reduced routine. At present, acquiring key data features, such as the number of household members and ambient temperature, is challenging, further complicating the modeling process.

After testing the abnormal time segment detection model for each user, the F1-score, our primary performance metric, displayed impressive results. However, deeper analysis revealed room for improvement in the Recall score. Moreover, missing values were commonly observed in the electricity consumption data for each residential user, often due to smart meter disconnections or maintenance. Ensuring comprehensive and accurate data for training could boost the model's performance. Beyond refining data quality and securing additional data features, we are optimistic about employing generative adversarial networks to generate electricity consumption behaviors more tailored to users, leveraging the data from the test dataset. This approach aims to address data insufficiencies and enhance the model's generalization capabilities. In simulating abnormal behavior, we simply multiplied past electricity consumption data by a random constant. Moving forward, we intend to investigate the model's efficacy under varied anomaly degrees in experimental setups, with the goal of further refining our proposed model's ability to detect abnormal electricity consumption time segments.

Acknowledgement

This research was partially supported by Bureau of Energy, Ministry of Economic Affairs, Taiwan (Grant No. 112-E0208).

References

[1] A. M. Soomro, G. Bharathy, N. Biloria, M. Prasad, A review on motivational nudges for enhancing building energy conservation behavior, *Journal of Smart Environments and Green Computing*, pp. 3–20, March, 2021. DOI: 10.20517/jsegc.2020.03

[2] M. A. Pimentel, D. A. Clifton, L. Clifton, L. Tarassenko, A review of novelty detection, *Signal Processing*, Vol. 99, pp. 215–249, June, 2014. https://doi.org/10.1016/j.sigpro.2013.12.026

[3] Y. Zhang, W.-W. Chen, J. Black, Anomaly detection in premise energy consumption data, *2011 IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, 2011, pp. 1–8. DOI: 10.1109/PES.2011.6039858

[4] J.-S. Chou, A. S. Telaga, Real-time detection of anomalous power consumption, *Renewable and Sustainable Energy Reviews*, Vol. 33, pp. 400–411, May, 2014.

[5] J. Luo, T. Hong, M. Yue, Real-time anomaly detection for very short-term load forecasting, *Journal of Modern Power Systems and Clean Energy*, Vol. 6, No. 2, pp. 235–243, March, 2018.

[6] P. Jokar, N. Arianpoo, V. C. M. Leung, Electricity theft detection in AMI using customers' consumption patterns, *IEEE Transactions on Smart Grid*, Vol. 7, No. 1, pp. 216–226, January, 2016.

[7] A. Pinceti, L. Sankar, O. Kosut, Load redistribution attack detection using machine learning: a data-driven approach, *2018 IEEE Power and Energy Society General Meeting*, Portland, OR, USA, 2018, pp. 1–5.

[8] C. Fan, F. Xiao, Y. Zhao, J.-Y. Wang, Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data, *Applied Energy*, Vol. 211, pp. 1123–1135, February, 2018.

[9] T. Śmiałkowski, A. Czyżewski, Detection of Anomalies in the Operation of a Road Lighting System Based on Data from Smart Electricity Meters, *Energies*, Vol. 15, No. 24, Article No. 9438, December, 2022.

[10] M. Carratù, V. Gallo, A. Pietrosanto, P. Sommella, G. Patrizi, A. Bartolini, L. Ciani, M. Catelani, F. Grasso, Anomaly Detection on Industrial Electrical Systems using Deep Learning, *2023 IEEE International Instrumentation and Measurement Technology Conference*, Kuala Lumpur, Malaysia, 2023, pp. 1–6.

[11] C.-W. Tsai, K.-C. Chiang, H.-Y. Hsieh, C.-W. Yang, J. Lin, Y.-C. Chang, Feature extraction of anomaly electricity usage behavior in residence using autoencoder, *Electronics*, Vol. 11, No. 9, Article No. 1450, May, 2022.

[12] F. Y. Edgeworth, Xli. On discordant observations, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. 23, No. 143, pp. 364–375, 1887.

[13] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, *ACM Computing Surveys*, TR-017, August, 2009.

[14] T. Akiba, S. Sano, T. Yanase. T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, Anchorage, AK, USA, 2019, pp. 2623–2631.

Biographies

Wen-Jen Ho received the Ph.D. degree in Electronic Engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 1998. He is currently a Senior Technical Director and the AI Engineering Technology Center Director at the Software Technology Institute within the Institute for Information Industry in Taiwan. His research interests include multimedia, embedded system, 3D/VR and AI.

Hsin-Yuan Hsieh is currently an undergraduate student at the School of Computer Science and Information Engineering at Taitung University. His research interests include anomaly detection, machine learning and Front-end technologies.

Chia-Wei Tsai received the Ph.D. degree in Computer Science and Information Engineering from National Cheng Kung University, Tainan, Taiwan, in 2011. He is currently an Associate Professor of National Taichung University of Science and Technology. His research interests include quantum cryptography, quantum information & computation, and anomaly detection.