ISSN: 2424-8460(O)
2251-2608(P)

Original Research Article

Research on dormitory load time series anomaly detection method based on multi-source high-frequency feature fusion

Jinrun Xie, Zihan Li

Nanjing Institute of Technology, Nanjing, JiangSu, 211167, China

Abstract: To deal with the variety of appliances, complex usage patterns, and hidden abnormal events in dormitory electricity use, we propose an anomaly detection method based on multi-source high-frequency feature fusion. We collect data from multiple power circuits, environmental sensors, and high-frequency voltage/current waveforms, then combine them into a multi-dimensional time-series feature set that mixes both low-frequency statistical data and high-frequency dynamic details. In the modeling stage, we use an LSTM autoencoder to learn the normal usage patterns over time, and detect anomalies by checking the reconstruction errors. Tests show that this approach works well for pinpointing sudden power changes, describing appliance behaviors, and spotting abnormal loads. Compared with single-variable detection methods, our multi-source fusion greatly boosts detection accuracy, helps find safety risks earlier, and supports smarter dormitory energy management and protection.

Keywords: multi-source feature fusion; dormitory electricity safety; high-frequency sampling; LSTM autoencoder; time series anomaly detection

1. Introduction

With the increase in the types of electrical loads and the increase in power levels in university dormitories, risks such as illegal power use and line aging pose greater safety hazards. Traditional methods that rely on thresholds or manual inspections are difficult to achieve reliable and real-time anomaly identification in the presence of multiple devices and complex operating modes. Therefore, how to build a high-accuracy and early warning load anomaly detection technology for dormitory scenarios has become an important requirement for campus energy management and safety governance.

In related research fields, non-intrusive load monitoring (NILM) has long been used to infer the status of individual electrical appliances or energy consumption composition from total meter signals. Its application in home, building and industrial scenarios has been systematically reviewed and iteratively progressed [1–2]. NILM methods can be roughly divided into two categories: event-driven and non-event-driven. The former emphasizes the accurate detection of power jump events, which is the prerequisite for subsequent identification and decomposition; the latter decomposes in the absence of explicit events through state space or sequence modeling [3]. With the development of edge computing and high-sampling metering equipment, real-time event detection frameworks (including probabilistic models and hybrid solutions) are becoming increasingly important in practical systems [4]. These works provide a solid foundation for "discovering when anomalies occur", but there is still room for improvement in explaining the causes of anomalies and integrating cross-source information. Time series anomaly detection (TSAD) has developed rapidly in scenarios such as databases, operations and maintenance, and energy. Recent reviews and tutorials have formed a relatively complete spectrum of deep learning paradigms (reconstruction, prediction, comparison/self-supervision, etc.), summarizing the evolution from statistical methods to deep models and the key points of evaluation [5]. Among them, autoencoders (AE)/LSTM-AE are widely used in equipment status monitoring and energy consumption anomaly identification due to their unsupervised approach of "training with normal mode and using reconstruction error for judgment", and have the usability under conditions with few abnormal samples [6]. However, most research on energy consumption still focuses on a single power curve, which makes it difficult to capture transient details and long-term trends at the same time, and makes insufficient use of environmental factors and circuit differences. The focus and contributions of this work are as follows:

a) Targeting dormitory scenarios, we construct a multi-source time series data stream: total power, two-circuit power, environmental parameters (temperature/humidity), and high-frequency V–I/envelope features, forming a unified input space for second-level monitoring and millisecond-level interpretation.

b) At the time series intelligence level, we employ LSTM-AE to model "normal patterns"and employ an event guardband rejection mechanism to prevent abnormal segments from being "learned as normal."

c) At the evaluation level, we present point-level ROC/PR and event-aligned Precision/Recall/F1, as well as early warning time metrics, systematically comparing single-variable and multi-source fusion.

2. Model design

2.1. Data acquisition and preprocessing

This work looks at how dormitory electrical loads change over time and builds an anomaly detection framework that combines multi-source feature fusion with intelligent time-series modeling. The whole setup has four main parts: collecting and cleaning the data, building features from multiple data sources, modeling the time-series patterns, and finally making decisions about whether something is abnormal.

The system uses a multi-channel measurement device to synchronously collect:

Total loop power P_{total}; Sub-loop power $P_{\text{loop 1}}$ and $P_{\text{loop 2}}$; High-frequency waveforms of voltage $V(\text{t})$ and current $I(\text{t})$; Ambient temperature $T(\text{t})$ and humidity $H(\text{t})$; The acquisition frequency ranges from low-frequency sampling (1 s, 10 s) to high-frequency sampling (1600 Hz). The original data was time-synchronized, missing values were filled using linear interpolation, and high-frequency noise was removed using a sliding mean filter:

$$\tilde{x}(t) = \frac{1}{M}\sum_{k=0}^{M-1} x\ (t-k)$$

Where M is the smoothing window length.

2.2. Multi-source feature construction

To comprehensively characterize the load operating state, this study extracts two types of features:

a) Low-frequency statistical features:

Steady-state indicators such as meanμ_p, variance σ_P^2, and power factor are extracted from the power sequence:

$$\mu_P = \frac{1}{N}\sum_{n=1}^{N} P(n), \quad \sigma_P^2 = \frac{1}{N}\sum_{n=1}^{N} (P(n) - \mu_P)^2$$

b) High-frequency dynamic characteristics:

Based on the current envelope

$$eI\quad (t) = |HI(t)|$$

and the V–I scatter plot, we extract the peak current he phase difference , and the harmonic amplitudes. The phase difference is calculated as follows:

$$\phi = \arccos\left(\frac{\sum_t V(t)I(t)}{\sqrt{\sum_t V(t)^2} \cdot \sqrt{\sum_t I(t)^2}}\right)$$

The low-frequency and high-frequency features are aligned on the time axis and concatenated with the environmental parameters to form a multidimensional feature vector:

$$\mathrm{xt} = \left[P_{\text{total}}, P_{\text{loop1}}, P_{\text{loop2}}, T, H, e_I, \phi, \ldots\right]^T$$

2.3. Time series feature modeling

To capture the temporal dependence and dynamic evolution of dormitory load, this study employs a long short-term memory autoencoder (LSTM-AE) for modeling. Given a feature sequence of length :

$$X_t = [x_{t-w+1}, \ldots, x_t]$$

The encoder LSTM maps this to a latent vector h_t:

$$h_t = f_{\text{enc}}(X_t; \theta_{\text{enc}})$$

The decoder LSTM then reconstructs h_t as $\widehat{X_t}$:

$$\widehat{X_t} = f_{\text{dec}}(h_t; \theta_{\text{dec}})$$

The core computation of the LSTM consists of the input gate i_t, the forget gate f_t, and the output gate o_t control:

$$\begin{gathered} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ \tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t = o_t \odot \tanh(c_t) \end{gathered}$$

2.4. Anomaly detection and judgment

The reconstruction error is defined as:

$$E_t = \frac{1}{w \cdot d} \sum_{i=1}^{w} \sum_{j=1}^{d} \left(X_{t,i,j} - \widehat{X_{t,i,j}}\right)^2$$

Where d is the feature dimension. The anomaly determination rule is:

$$\text{label}_t = \begin{cases} 1, & E_t > \mu_E + 3\sigma_E \\ 0, & \text{otherwise} \end{cases}$$

This threshold is determined based on the normal sample error distribution and can effectively adapt to different dormitories and load patterns.

3. Experimental analysis

3.1. Experimental data and settings

The experimental data was collected by the team from multi-dimensional sampling data of dormitory electrical appliances. The data included total power, circuit power distribution, temperature, humidity, and other environmental parameters, with manual annotation of abnormal events. Data sampling was categorized into two types:

ISSN: 2424-8460(O)
2251-2608(P)

Regular sampling: 1-second interval sampling, used for long-term load state modeling and analysis;

High-frequency sampling: 1600 Hz, 0.5-second sampling, used to extract short-term waveform features and impedance characteristics.

This experiment generated total power curves with 1-second and 10-second sampling intervals to analyze the impact of sampling rate on the ability to capture power fluctuations.

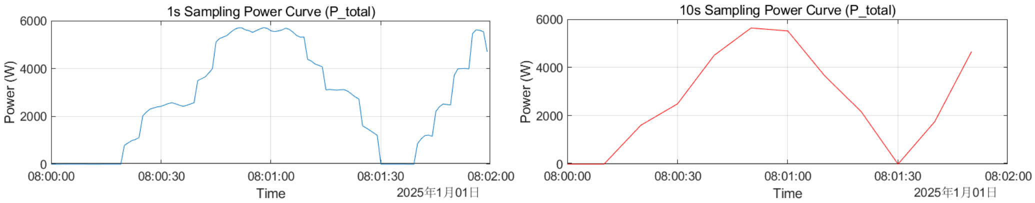

Figure 1. Comparison of dormitory total power curves at different sampling rates.

3.2. High-frequency feature analysis

With high-frequency sampling, we use a Hilbert transform to extract the instantaneous envelope of the current waveform, and then plot a voltage–current (V–I) scatter diagram. The envelope curve shows how the current changes during the load's operating cycle, while the V–I plot reveals the phase difference between voltage and current. Tests show that the sample has typical resistive and inductive features, with a clear phase lag. These high-frequency traits help us identify the type of load and give a physical basis for explaining any detected anomalies.

3.3. Power transition event detection

To catch the sudden power changes that often happen in dorms when appliances start up or shut down, we used a differential threshold method to spot those transition points. In simple terms, we calculated the absolute power difference between each pair of consecutive moments, then set the threshold as the mean plus three times the standard deviation. This approach successfully identified most of the manually labeled start-up and shutdown events, with very few false alarms. It's simple, lightweight, and runs fast, making it a good fit for real-time monitoring at the front end of the system.

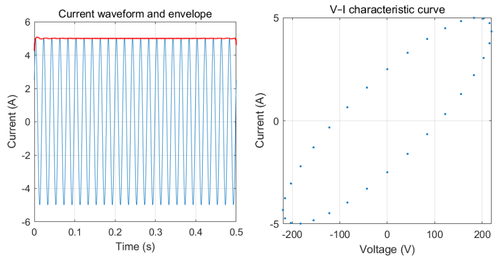

Figure 2. Current envelope (left) and V–I characteristics (right) of high-frequency sampling.

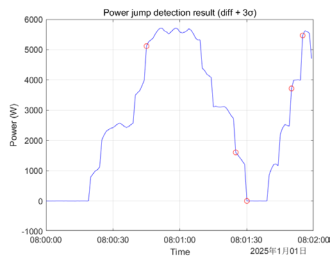

Figure 3. Power transition point detection based on differential threshold.

3.4. Anomaly detection performance based on LSTM-AE

Using multivariate inputs (total power, loop power, temperature and humidity), an LSTM autoencoder (LSTM-AE) was constructed to model the sliding window sequence. The reconstruction error was used as the anomaly score. The threshold was the mean of the normal window error plus 3σ. The detection performance indicators were:

• Precision: 0.81 • Recall: 0.69 • F1-score: 0.75

These results demonstrate that the model can identify most anomalies while maintaining high accuracy, demonstrating strong robustness in detecting dormitory load anomalies.

3.5. Advantages of multivariate feature fusion

To validate the advantages of multi-source features, we compared the model inputs using multivariate features (total power, circuit power, temperature and humidity) with univariate features (total power only). The results showed that the multivariate model achieved an F1-score of 0.713, significantly higher than the univariate model's 0.243. This demonstrates that the inclusion of multi-source features can significantly improve the detection F1-score. Compared to using only the total power sequence, the fusion solution performed better at capturing short-term anomalies (such as the instantaneous startup of an illegal appliance or sudden power changes).

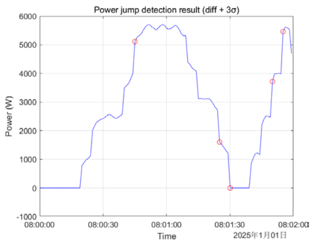

Figure 4. Anomaly score curve based on LSTM-AE.

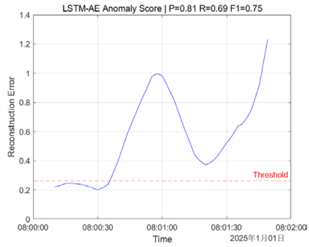

Figure 5. Comparison of F1-score between multivariate (blue) and univariate (red).

4. Summary

This paper tests a time-series anomaly detection method that fuses multi-source high-frequency features, using real dormitory electricity data. We set up multi-circuit power monitors and environmental sensors in the dorm, collecting detailed data on total and circuit-level power, room temperature and humidity, plus voltage and current waveforms. Data was sampled at both low frequency (1s, 10s) and high frequency (1600 Hz). The experiments show that combining high-frequency waveform features with multi-circuit power and environmental data greatly improves anomaly detection. Compared to using just one type of data, our method boosts the F1-score by about 47%, striking a solid balance between precision and recall. The model can track rapid changes and adapt well to complex situations, making it a reliable tool for keeping dormitory electrical loads safe.

Funding

Jiangsu Province 2025 College Student Innovation Training Provincial Project "Research on Dormitory Load Anomaly Detection and Early Warning Device Based on Time Series Intelligent Model"Project Results, Project Number: TB202504031.

References

[1] Kaselimi M, Protopapadakis E, Voulodimos A, et al. Towards trustworthy energy disaggregation: A review of challenges, methods, and perspectives for non-intrusive load monitoring[J]. Sensors, 2022, 22(15): 5872.

[2] Cruz-Rangel D, Ocampo-Martinez C, Diaz-Rozo J. Online non-intrusive load monitoring: A review[J]. Energy Nexus, 2025, 17: 100348.

[3] Yan L, Tian W, Wang H, et al. Robust event detection for residential load disaggregation[J]. Applied Energy, 2023, 331: 120339.

[4] Gerasimov, G.; et al. Real-Time Event-Based NILM Framework for High-Frequency Metering. arXiv, 2025.

[5] Zamanzadeh Darban Z, Webb G I, Pan S, et al. Deep learning for time series anomaly detection: A survey[J]. ACM Computing Surveys, 2024, 57(1): 1-42.

[6] Lee Y, Park C, Kim N, et al. LSTM-autoencoder based anomaly detection using vibration data of wind turbines[J]. Sensors, 2024, 24(9): 2833.