Energy Proceedings
Vol 54, 2025
ISSN 2004-2965

Autoregressive Transformer for Predicting and Synthesizing Residential Load#

Kosaku Nakano[1], Daishi Sagawa[1], Shingo Ito[2], Keisuke Ando[2], Ryohta Kitagawa[2], Kenji Tanaka[1*]

1 Graduate School of Engineering, The University of Tokyo

2 Rezil Inc.

(Corresponding Author: tanaka@tmi.t.u-tokyo.ac.jp)

ABSTRACT

Accurate load forecasting is crucial for efficient energy management, particularly due to the increasing adoption of renewable energy sources and the growing need for grid stability. This study introduces an autoregressive transformer encoder-decoder model that advances residential electricity load prediction and synthesis. Unlike the conventional models, our proposed model utilizes temperature and calendar information as primary inputs and can generate an electricity load without relying on the past load. This model's design is rooted in the observation that temperature and calendar context sufficiently capture residential electricity consumption dynamics. By excluding the reliance on past load data, our model is able to synthesize electricity load for arbitrary temperature and calendar scenarios in addition to effectively predicting the future load. This synthesizing capability is invaluable for optimizing the planning and operation of residential energy systems, including photovoltaic systems and batteries. Experimental results using real-world electricity load data of residential buildings demonstrate the effectiveness of our transformer-based model, offering a robust framework for future load prediction and synthesis under varying conditions.

Keywords: Electricity Load Forecasting, Electricity Load Synthesis, Transformer, Residential Electricity Load, Energy Management Systems

1. INTRODUCTION

Accurate load forecasting is crucial for efficient energy management [1]. With the proliferation of renewable energy and the push towards decarbonization, the uncertainty in energy supply has increased. As a result, accurately predicting the balance between demand and supply to maintain the stability and reliability of the power grid has become even more critical. In the residential sector, electricity load varies due to factors such as seasons, weather, and calendar events (e.g., weekdays and holidays) [2, 3, 4]. Therefore, precise load forecasting that considers these factors is essential for effective energy management.

Particularly in recent years, there has been an increase in residential buildings equipped with photovoltaic (PV) systems and batteries [5]. For these buildings, accurate load forecasting can optimize the costs associated with PV and battery installations, providing significant economic benefits to consumers. Moreover, achieving precise load forecasting could encourage the adoption of PV systems and batteries in residential buildings that do not yet have them. This would benefit the overall energy system by increasing the share of renewable energy sources and enhancing the system's capacity to absorb these sources, leading to greater sustainability and efficiency.

Furthermore, synthesizing load data for specific buildings is essential for optimal planning of PV and battery sizes. By generating accurate load profiles based on various temperature and calendar scenarios, energy planners can make informed decisions about the most efficient and cost-effective sizes for PV systems and batteries. This tailored approach ensures that the energy needs of the building are met while maximizing the economic and environmental benefits. Accurate synthesized load data can also help evaluate different scenarios and plan for peak demand periods, ultimately leading to more resilient and sustainable energy systems.

Because electricity load forecasting is of paramount importance for residential energy management, a significant amount of research has been conducted in this area. Various models and methodologies have been developed, leveraging data such as historical load records, weather forecasts, and socio-economic factors to enhance prediction accuracy. Some studies have employed classical methods such as ARIMA [6], while others have used regression trees for high-precision forecasts [7]. More recent approaches include

This is a paper for the 16th International Conference on Applied Energy (ICAE2024), Sep. 1-5, 2024, Niigata, Japan.

probabilistic forecasting using LSTM [8], and with the advancement of self-attention mechanisms, transformer-based models have also been explored [9]. However, most existing models typically use past load data as inputs. Considering that residential electricity load is strongly influenced by weather and calendar information, such as weekdays and holidays, rather than by past loads, this approach may have limitations in accurately predicting future consumption.

We observed that residential building loads are predominantly influenced by temperature and calendar data. This observation led us to conceive of residential buildings as a model that takes temperature and calendar information as inputs and outputs electricity load. In this study, we propose an autoregressive transformer model that predicts future load based solely on temperature and calendar information, without relying on past load data. We chose the Transformer encoder-decoder primarily because it has demonstrated high performance in other modalities, such as text and images [10, 11]. This model is able to accurately predict future loads and can also generate loads for all buildings in the dataset, using the same model for each building. Furthermore, due to its design, the model can synthesize residential load for arbitrarily given temperature and calendar scenarios. This capability is particularly valuable for optimizing PV and battery sizes and enhancing overall energy system resilience.

2. MATERIAL AND METHODS

2.1 Time-Series Prediction and Synthesis

This study proposes a time-series prediction or synthesis not relying on a past sequence. Fig. 1 illustrates the distinction between the conventional approach and our proposed approach. On the left side, the conventional approach is depicted. Here, the model predicts future target values, denoted as $\boldsymbol{x}_{t+1:t+T}$, using both past values $\boldsymbol{x}_{1:t}$ and relevant covariates $\boldsymbol{s}_{1:\mathrm{t+T}}$. On the right side, our proposed approach is shown. In this approach, the model generates future values $\boldsymbol{x}_{t+1:t+T}$ based on the condition $\boldsymbol{c}_t$. $\boldsymbol{c}_t$ is data or information available at time t and can include $\boldsymbol{s}_{1:\mathrm{t+T}}$.

2.2 Transformer Encoder-Decoder-Based Model

We propose a Transformer encoder-decoder-based model to predict and synthesize daily residential load profiles, as depicted in Fig. 2. The model, capable of handling any timeframe, focuses on a three-day target ($T = 3$) for this study.

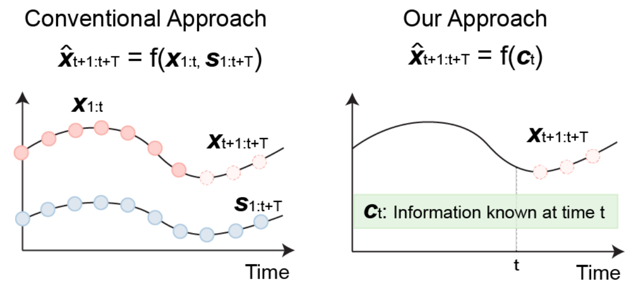

Fig. 1 Comparison between the conventional approach and our approach.

The model takes several inputs for each residential building: a building ID, a seven-day temperature history, a three-day temperature forecast of target days, and information on weekdays and holidays for the target days. These inputs are transformed into latent representation $\boldsymbol{z_c}$, where the temperature sequence is processed through a linear layer and other data through lookup table embeddings. The Transformer encoder processes $\boldsymbol{z_c}$ to generate a latent representation $\boldsymbol{z_{enc}}$.

The decoder employs causal self-attention, generating daily load data for the target days from learnable <start of sequence> (<sos>) tokens in an autoregressive manner. This <sos> token can optionally be replaced by a context token, which is a latent representation of past loads. The latent representation output from the decoder is mapped to the target load by another linear layer. The cross-attention layers within the decoder compute attention between the encoder outputs $\boldsymbol{z_{enc}}$ and the generated sequences, facilitating accurate load prediction based on the input conditions.

The Transformer architecture comprises two subnetworks in both the encoder and decoder: self-attention and feed-forward. An additional subnetwork in the decoder includes cross-attention, enhancing interaction between encoder outputs and decoder predictions. Each sublayer starts with layer normalization (pre-norm architecture) and has a residual connection.

The model operates with a low-dimensional latent space (dimension=32) to enhance learning efficiency, configured with 6 layers and 8 attention heads in both the encoder and decoder.

2.3 Dataset

In this study, we utilize half-hourly load data collected from more than 400 residential buildings in Japan over a period of about two years. The sizes of the residential buildings vary, and so do their load magnitudes. Weather forecast data and actual weather

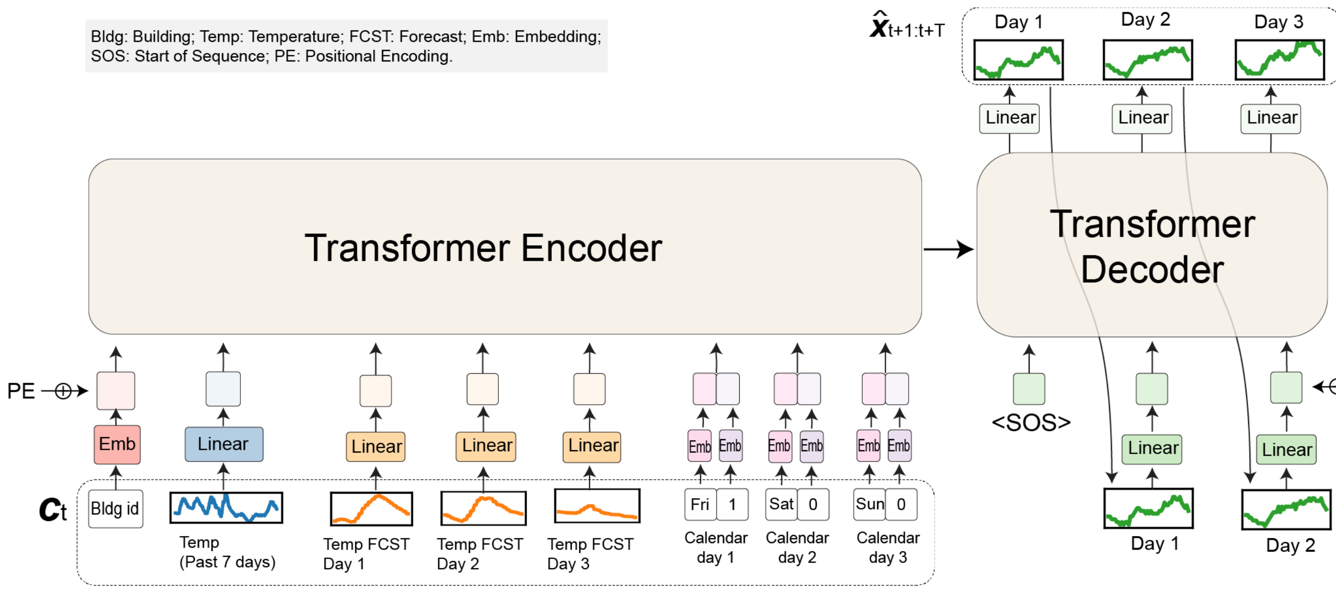

Fig. 2 Overview of our proposed method.

data were also collected. Together with the load data, these form the dataset used in this work. Data from the last 30-day period in the dataset are designated as the test split, while the remainder is used for training.

2.4 *Model Training and Evaluation*

For the model training process, mean squared error (MSE) is employed as a loss function. The model is trained for a total of 500,000 steps with a batch size of 256. Adam is employed as an optimizer with a learning rate of 8e-4.

To stabilize the training and improve the model's performance, the temperature and load data are normalized across the entire training dataset using the maximum and minimum values. The same maximum and minimum values are used during testing to normalize and de-normalize inputs and outputs.

3. RESULTS AND DISCUSSION

We evaluated the trained model's prediction accuracy using approximately 12,000 test instances, achieving a root mean squared error (RMSE) of 0.0076 and a mean absolute error (MAE) of 0.0045 on normalized data, demonstrating high precision in the forecasts.

To further assess the predictive capability of our model, randomly selected qualitative results are presented in Fig. 3. While some instances reveal challenges in accurately predicting troughs and peaks in the loads (as seen in the top left and top third from left plots), other examples successfully capture complex load

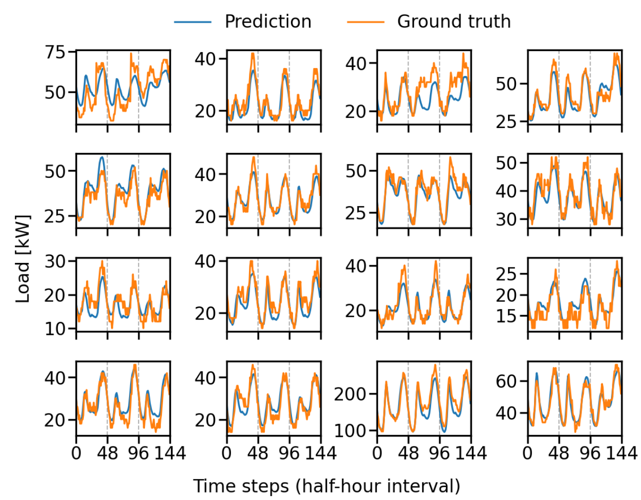

Fig. 3 Visualization of Prediction Results.

patterns, accurately reflecting both the troughs and peaks. It is noteworthy that the plot in the bottom right illustrates how the model accurately predicts the load on the third day (time step >= 96), where the load pattern deviates from the previous days. Moreover, the proposed model handles a wide range of load scales, from as low as 10 kW to over 200 kW, as shown in Fig. 3.

To evaluate the model's ability to synthesize load profiles, we conducted an experiment where the model generated load output for a building under various temperature conditions. The model synthesized the load for five different temperature scenarios, each with an average temperature over three days: 1.28°C, 6.28°C, 14.28°C, 22.28°C, and 38.28°C. These temperature

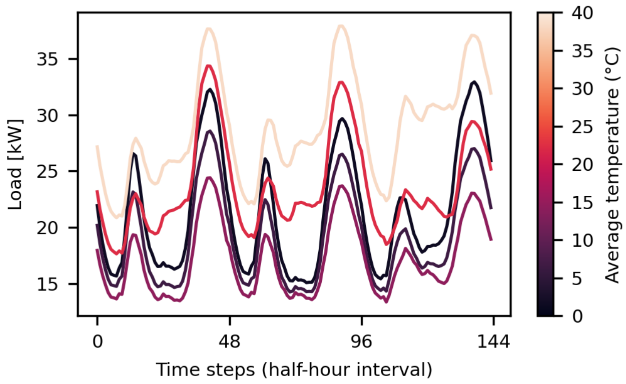

Fig. 4 Results of load synthesis for different temperature scenarios.

profiles were created by adding constant values to a temperature profile sampled from the test dataset. Fig. 4 presents the results of the synthesis. Across all temperature scenarios, the model successfully captures typical residential load patterns, characterized by an increase in load during the morning, a slight decrease during the day, and a peak in the evening and night.

Additionally, the model effectively reflects variations in load patterns across different temperature scenarios. For the moderate temperature scenario (14.28°C), the load remains the lowest among the five scenarios throughout the day. In contrast, both high and low-temperature scenarios (38.28°C and 1.28°C, respectively) show higher loads. The load during the high-temperature scenario remains elevated throughout the day, whereas the load in the low-temperature scenario follows a distinct pattern, with higher demand in the morning and evening and lower demand during the daytime. These differing patterns reflect varying heating and air conditioning usage based on temperature conditions.

4. CONCLUSION

This study demonstrates the effectiveness of an autoregressive Transformer-based model that does not rely on past load data in generating residential load profiles. Utilizing data from more than 400 residential buildings over about two years, the model showed high accuracy and stability for predicting loads. Importantly, the model's design also enables the synthesis of load profiles under arbitrary temperature and calendar conditions. Despite some challenges in predicting troughs and peaks in several cases, the overall performance metrics, RMSE and MAE, confirm the model's robustness. Future efforts will focus on improving the predictive performance and further exploring the model's potential for synthesizing load profiles, thus extending its applicability.

REFERENCE

[1] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustainable Cities and Society. 2020;55:102052.

[2] Fikru MG, Gautier L. The impact of weather variation on energy consumption in residential houses. Applied Energy. 2015 Apr 15;144:19-30.

[3] Lam JC, Tang HL, Li DH. Seasonal variations in residential and commercial sector electricity consumption in Hong Kong. Energy. 2008 Mar 1;33(3):513-23.

[4] Wang H, Alattas KA, Mohammadzadeh A, Sabzalian MH, Aly AA, Mosavi A. Comprehensive review of load forecasting with emphasis on intelligent computing approaches. Energy Reports. 2022 Nov 1;8:13189-98.

[5] Reimuth A, Locherer V, Danner M, Mauser W. How Does the Rate of Photovoltaic Installations and Coupled Batteries Affect Regional Energy Balancing and Self-Consumption of Residential Buildings?. Energies. 2020 May 29;13(11):2738.

[6] Cho MY, Hwang JC, Chen CS. Customer short term load forecasting by using ARIMA transfer function model. In Proceedings 1995 International Conference on Energy Management and Power Delivery EMPD'95 1995 (Vol. 1, pp. 317-322). IEEE.

[7] Lusis P, Khalilpour KR, Andrew L, Liebman A. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. Applied Energy. 2017;205:654-69.

[8] Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C. Probabilistic individual load forecasting using pinball loss guided LSTM. Applied Energy. 2019;235:10-20.

[9] Zhao H, Wu Y, Ma L, Pan S. Spatial and temporal attention-enabled transformer network for multivariate short-term residential load forecasting. IEEE Transactions on Instrumentation and Measurement. 2023.

[10] Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research. 2020;21(140):1-67.

[11] Yu J, Xu Y, Koh JY, Luong T, Baid G, Wang Z, Vasudevan V, Ku A, Yang Y, Ayan BK, Hutchinson B. Scaling autoregressive models for content-rich text-to-image generation. arXiv preprint arXiv:2206.10789. 2022;2(3):5.