E3S Web of Conferences **522**, 01018 (2024) https://doi.org/10.1051/e3sconf/202452201018
VESEP2023

Response load prediction of demand response users based on parallel CNN

Haowen Sun[1], *Mei* Yu[1,*], *Bitao* Li[1], *Liping* Fan[1,2], *Jie* Yu[1,2], *Dong* Yu[1,2] , and *Shouzhi* Xu[1]

[1]Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang, Hubei, China
[2]Yichang Power Supply Company of State Grid Hubei Electric Power Co., Ltd., Yichang, Hubei, China

Abstract. YAs China advances its transition towards green and low-carbon energy, the proportion of new energy generation in the power grid is gradually increasing, leading to a significant rise in the demand for power resource scheduling. However, due to the scarcity of historical load response data from users, it is challenging to effectively predict user-responsive loads. To address this issue, this study proposes a method of augmenting historical load response data in a weakly supervised manner. Taking into account the unique circumstances of high-voltage users, a sparse CNN for anomaly detection is introduced, along with a multi-branch parallel CNN model capable of weighted output of prediction results from both global and local perspectives. Subsequently, effective iterative training of the model is performed using the EM algorithm. Ultimately, accurate prediction of user-responsive loads is achieved. Based on historical 96-point load data and load response data from high-voltage users in a specific city in China, the predicted results are compared with actual load response data, validating the rationality and accuracy of this method in predicting user-responsive loads.

1 Introduction

As China advances towards green and low-carbon energy transformation and development, striving for carbon neutrality, the proportion of new energy sources in the power grid continues to rise. This trend has led to a notable surge in the demand for grid flexibility and adjustment resources. Among these resources, user-side demand response (DR) plays a pivotal role in alleviating pressure during peak periods of power demand.

In today's landscape, numerous deep learning models are employed for time series load forecasting in the realm of electricity consumption. These models primarily include LSTM, CNN, Transformer, and others. However, when predicting electricity load, these models heavily rely on a substantial volume of historical electricity data for effective training. Unfortunately, due to the limited availability of demand response data, particularly historical response load data, the training of models for responsive load prediction faces significant challenges. Consequently, the predictive accuracy of responsive load values suffers.

* Corresponding author: yumei@ctgu.edu.cn

In order to fill the technical gap in demand response load forecasting, this paper endeavors to train a commonly used model for load forecasting based on historical response load data to predict user response load. However, due to the scarcity of historical response load data, anomalies occur in historical electricity load data due to holidays and other circumstances. Consequently, it is observed that user response load forecasting differs from traditional user power load forecasting. Responsive load forecasting lacks a temporal dependency similar to that in power load forecasting models. Therefore, forecasting responsive load can be challenging, as it relies on direct training with existing data and models. This presents difficulties, and the prediction results may not be ideal.

Based on the aforementioned challenges, this paper initially conducts thorough data preprocessing on historical response load data and historical electricity load data. Subsequently, it proposes a Parallel CNN model, comprising several locally-viewed parallel sparse CNNs alongside a global perspective CNN. This model utilizes a weighted output to forecast both local and global field of view results. By doing so, the model aims to mitigate the impact of user response during holidays or production shutdowns on load prediction. Given the limited historical response load data, the basic model trained may lack robustness. Hence, a weak supervision technique is employed to augment the historical response load data. Initially, a fundamental Parallel CNN model is trained using actual response load data. Subsequently, this model labels the load data, generating a larger-scale dataset. This iterative process is repeated multiple times, adhering to the EM paradigm to ensure the reliability of the acquired data.

The prediction accuracy of this model surpasses that of traditional load forecasting models. Not only does it address the challenge of ineffective model training due to the scarcity of response load data, but it also enables more precise prediction of user response load, particularly during holiday or production shutdown periods.

2 Related work

With the advancement of deep learning, significant progress has been made in electric load forecasting. Utilizing these technologies, researchers have developed various deep learning models tailored for this task. These models learn from historical data, accurately predicting future electricity consumption. Neural networks, notably MLP[1], LSTM, CNN[2], and CNN-LSTM, enhance flexibility and efficiency, capturing intricate patterns for precise load change modeling, thus improving forecasting accuracy and efficiency.

2.1 LSTM and CNN in load forecasting

In contemporary times, a plethora of innovative load forecasting models have emerged, harnessing the capabilities of CNN, LSTM, and CNN-LSTM architectures.

The MTMV-CNN-LSTM[3] deep learning framework for electricity load forecasting integrates multi-task learning, CNN, and LSTM to enhance both generalization and accuracy, presenting multi-task learning as a solution to the intricate nature of weather-related factors. In ultra-short-term demand forecasting for industrial power loads, a hybrid ensemble learning model is proposed, utilizing LSTM networks[4]. It integrates bagging, random subspace, and boosting into its ensemble strategy, along with a novel loss function that incorporates peak demand prediction errors to address bias-variance trade-off. The model demonstrates outstanding accuracy with minimal prediction errors across different metrics.

A novel dual-stage attention LSTM network for short-term regional load probability forecasting[5], with advancements in multi-source data utilization and generalization. Leveraging feature and temporal attention mechanisms, the model dynamically selects key features and explores temporal dependencies[6] proposes a novel method for short-term power load forecasting, utilizing CNN, LSTM, and attention mechanism to address information loss from overly long input time series data. The model employs a one-dimensional CNN layer for feature extraction, an LSTM layer for capturing temporal correlations, and introduces an attention mechanism to optimize the LSTM output weight, thereby enhancing key information influence and overall prediction model optimization.

2.2 Transformer in load forecasting

The literature[7] introduces a tailored deep learning prediction module based on CNNs-Transformer. This module constructs a deep learning framework consisting of three layers of CNN and Transformer, utilizing QRLoss for probability interval prediction across various confidence levels. Through the application of loss penalty techniques, the model's focus on weekends is enhanced, enabling superior performance in day-ahead power load forecasting on both weekdays and weekends. An enhanced ITFT model[8] is proposed for probabilistic forecasting of hourly load time series, addressing challenges posed by renewable energy growth and demand-side responsiveness. It replaces LSTM with GRU for more efficient long-term dependence learning.

The CEEMDAN-SE-TR hybrid model[9], proposed for short-term load forecasting, utilizes the Transformer (TR) model's attention mechanism to mitigate long memory loss issues, thereby achieving superior forecasting accuracy compared to other machine learning models.

3 Method

3.1 Problem statement

The prediction of user-responsive load belongs to a similar load prediction problem, it is different from the traditional electricity load prediction. This article sets the data set input into the model to $X = [x_0, x_1, x_2, \dots, x_{t-1}], (X \in R^{m \times n}, x_i \in R^n)$, Where x_t is the response day, X is the dataset containing the daily 96-point electricity load data of the responsive users up to t days before the response day, and x_i represents the daily 96-point electricity load data of the responsive user. The target set of the model is defined as $Y = [y_t], (y_t \in R^m)$, where y_t represents the response value of the responsive users on the response day.

3.2 Method statement

In this experiment, considering scenarios such as holidays, rest days, or routine shutdowns for maintenance in which participating users may not respond, certain regular missing values and distinctive anomalies appear in the daily 96-point electricity load data. To address this, a Parallel model consisting of four CNNs for processing local electricity load data and one CNN for processing global electricity load data was employed. The results obtained from the local and global models were weighted and combined for output, with loss calculated using the Mean Squared Error (MSE) function.

However, due to the infrequent occurrences of demand response events in routine life and production, resulting in limited actual response load data, the trained basic model lacked robustness and could not be widely applied in real-world scenarios. To tackle this issue of insufficient response data, a weak supervision approach for reasonable data augmentation was adopted.

Specifically, a basic model was trained based on real response load data, and this model was used to label the daily 96-point electricity load data. Subsequently, a larger-scale dataset was generated through iterative rounds of this process, following the Expectation-Maximization (EM) paradigm.

3.2.1 Parallel CNN architecture diagram

The proposed Parallel CNN model in this paper is designed to address special electricity consumption scenarios such as holidays or shutdowns for high-pressure users. It consists of multiple parallel local-time-step CNN models, each coupled with a global-data-input CNN model. The training dataset is structured as (batch_size, 100, 96) where batch_size represents the length of extracted data.

Within the local field of view, the data for each local CNN is fed with a time step covering the preceding 100 days, while the global CNN takes in the entire 100-day electricity consumption data. Subsequently, the predicted results from the local CNN models are averaged and combined with the output of the global CNN model, forming a weighted output. This process, illustrated in Fig. 1, ultimately yields the prediction results.

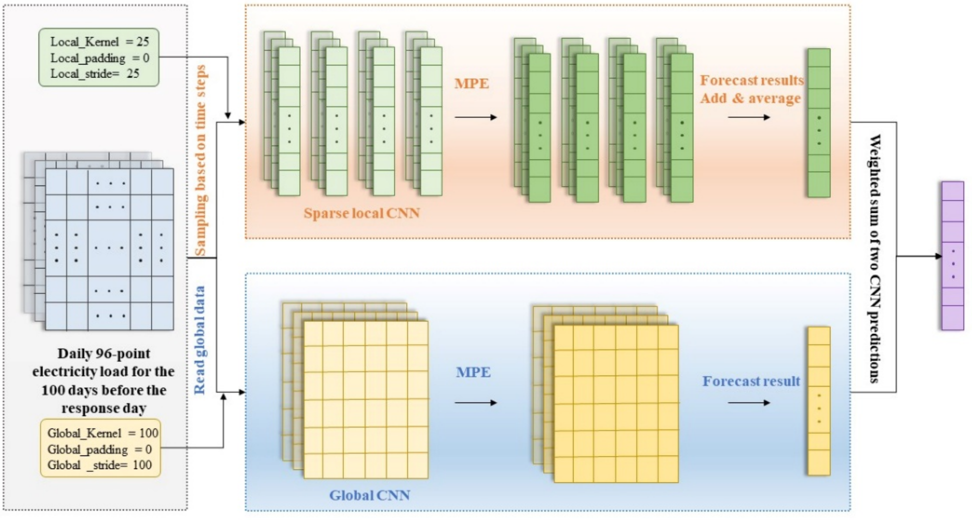

Fig. 1. Parallel CNN model.

The designed model was built using Pytorch, an open source neural network library written in Python. In the proposed model framework, the following hyperparameters are used to train the network.

(1) Convolution type: one-dimensional convolution.
(2) Convolution kernel size: local field of view is 25, global field of view is 100.
(3) Activation function: MPE function.
(4) Optimizer: Adam.

3.2.2 EM algorithm

The EM algorithm [10] is an iterative method utilized for maximum likelihood estimation of probabilistic model parameters that contain latent variables. In this experiment, owing to stable power supply in daily production and life, demand response events are infrequent, leading to a small number of users participating in the response. Consequently, the available historical response load data may not suffice to train a suitable response data prediction model.

As a solution, a rudimentary model is initially trained with existing real response load data. Subsequently, the load data is labeled through the model to generate a larger volume of response data. This process is iterated over multiple rounds, adhering to the principles of the EM algorithm.

The specific steps of the algorithm are as follows in Table 1:

Table 1. Weakly-Supervised EM algorithm.

Algorithm1 Weakly-Supervised EM
Input: CNN initialization parameters $\theta^{(0)}$, Historical response user load data $x_{small} \in R^{l\times d}$, Unresponsive user's electricity load data $x_{large} \in R^{l\times d}$, Historical response load data tags $z_{small} \in R^{l\times d}$, Unresponsive users can respond to load data tags $z_{all} \in R^{l\times d}$
Initial-Step:
1. $Z_{small}^{(0)} = V\left(Z_{small}^{(0)} \middle| Z_{small}^{(0)}\right);\ \theta^{(0)}$
2. $\theta_{small}^{(i)} = arg\max_{\theta} \sum_i \left(\frac{\hat{z}_{small} - z_{small}^{(i)}}{z_{small}^{(i)}}\right)$
E-Step:
$Z_{large}^{(i)} = M\left(x_{large};\ \theta_{small}\right)$
M-Step:
1. $Z_{all}^{(i)} = Z_{small}^{(i)} + Z_{large}^{(i)}$
2. $\theta_{all}^{(i)} = arg\max_{\theta} \sum_i \left(\frac{\hat{z}_{all} - z_{all}^{(i)}}{z_{all}^{(i)}}\right)$

Initial-Step: Initialize the parameters $\theta^{(0)}$ and train a preliminary model using historical power load data $x_{small} \in R^{l\times d}$ to derive the corresponding historical response load labels $z_{small} \in R^{l\times d}$. Then, proceed to calculate $\theta_{small}^{(i)}$ based on this training.

E-Step: Determine $Z_{large}^{(i)}$, which involves formulating the appropriate response label based on the power load data x_{large} from other users, utilizing θ_{small}.

M-Step: Utilize the $Z_{large}^{(i)}$ acquired in the E step to generate labels $Z_{all}^{(i)}$ for all samples, and retrain the model based on these updated labels to enhance its data fitting capabilities. Subsequently, leverage the dataset with these revised labels to train a fresh model.

In each iteration, the E-Step and the M-Step are alternately executed until convergence. The enhancement of responsive load data is achieved through a weak supervision method. This process adheres to the EM paradigm, ensuring that the predictions of the user-responsive load prediction model obtained through iteration are reasonable and accurate.

3.2.3 Sparse local CNN

Sparse local CNN have demonstrated unique advantages in handling power load data with local anomalies. In contrast to conventional CNNs covering all historical electricity load data, sparse CNNs employ a clever strategy to process the data from a local perspective through multiple parallel models. This effectively balances attention between normal and abnormal data.

While the traditional convolutional neural network method processes all historical data to capture the changing trend of electricity load comprehensively, it may encounter unnecessary interference due to abnormal conditions. The innovation of sparse convolutional neural networks lies in dividing the overall dataset into local fragments, denoted as $Z =$

$\{z_1, z_1, \ldots, z_n\}\left(n = \frac{days}{l}\right)$ using an appropriate step l. Here, the training centers the response t days before the current day. Each sparse CNN reads the time interval step size consistently, which is divisible by days. These local data fragments are then separately processed through multiple parallel models. This segmentation and parallel processing strategy enables a local view analysis of the data and effectively mitigates the impact of abnormal data on the overall model. Four parallel sparse convolutional neural networks simultaneously divide the same time period into four intervals of data at identical time steps. Consequently, the model focuses more on learning local features during the training process, thereby partially decoupling the relationship between normal and abnormal data.

3.2.4 Loss value

The Parallel CNN incorporates several local view CNNs alongside a global view CNN simultaneously. The local view CNN focuses on capturing the local features of the electrical load data, whereas the global view CNN concentrates on understanding the overall data distribution. To derive the final prediction result, it's necessary to combine and weigh the prediction results obtained from these two perspectives. Equation (1) is presented as follows:

$$y_{pred} = \omega_{local} \times \frac{\sum_{i=1}^{n} y_{local,i}}{n} + \omega_{global} \times y_{global} \tag{1}$$

The predicted values from the local view are averaged, and this average is then weighted and combined with the prediction from the global view to yield the final predicted value, denoted as y_{pred}. ω_{local} and ω_{global} represent the weights assigned to the average predicted value in the local field of view and the predicted value in the global field of view, respectively. These weights can be adjusted to improve the prediction effectiveness through subsequent ablation experiments.

Given that the prediction problem involves regression and accounts for variations in electricity usage among different users, the Mean Percentage Error (MPE) is utilized. This metric normalizes the disparities before computing squared differences, effectively mitigating the influence of label value scales on loss. As a result, the accuracy of model predictions becomes more intuitively discernible. Equation (2) is expressed as follows:

$$Loss = \frac{1}{N}\sum_{k=1}^{N}\left(\frac{\left|y_{pred}^{k} - y_{true}^{k}\right|}{y_{true}^{k}}\right) \tag{2}$$

In the equation, N represents the total number of samples, k denotes the k-th sample, and the mean square error signifies the average of the squared differences employed to assess the model's output against the true label.

4 Experiment

This paper addresses the issue of multiple continuous outliers in historical power consumption data among high-voltage users, attributed to factors such as line maintenance, holidays, or production shutdowns. To predict the responsive load during demand response periods, a Parallel sparse CNN model is designed.

Initially, common load prediction models are adapted to forecast responsive load. Remarkably, only the CNN model demonstrates convergence of the loss function. Building upon this, a Parallel sparse CNN model is crafted, comprising a CNN model for the global field of view and CNN models for four local fields of view, forming a parallel relationship. Additionally, the CNN models within each local field of view operate in parallel. Predictions from the four local CNNs are averaged and juxtaposed against the CNN model for the global field of view. The weighted output is consequently obtained as the model's prediction. Subsequently, due to limited historical response load data, a weak supervision method is employed for data augmentation. Initially, a basic Parallel CNN model is trained with actual response load data. This model labels the load data, generating augmented data with increased magnitudes. Multiple iterations are conducted, adhering to the Expectation-Maximization (EM) paradigm, ensuring the reliability of the obtained data throughout the process.

4.1 Data preprocessing

This article focuses on historical response load data and daily 96-point load data of users in a specific region of China as the research subject. To construct a reliable model, the training set comprises historical daily 96-point electricity load data of responding users from the 100 days preceding the response day, while the target set consists of historical response load data of responding users on the response day.

However, due to issues within the smart meter network, the electricity load data uploaded by users at 96 points in a single day contain irregular intermittent 0 values, which could detrimentally affect the model's predictive efficacy. Therefore, it is essential to perform prudent data preprocessing. Upon processing the user power load data and response load data, several anomalies were identified. In response to these anomalies, the following preprocessing measures were undertaken:

(1) If the power consumption data for an entire day consists entirely of 0 values for more than a week, the corresponding power consumption sample is deleted, as this likely indicates an abnormal situation.

(2) If the power consumption data for an entire day is entirely composed of 0 values, and the duration of such occurrences is less than one week, the day's electricity consumption data is deleted and replaced with the corresponding number of electricity consumption day data.

(3) For a single day of electricity consumption, if the number of consecutive 0 record points exceeds or equals half of the total, the data for that day is deleted, and the subsequent day's data is utilized. Conversely, if the number of consecutive 0 record points is less than half, the first 0 value point is replaced by the previous point with data, and the last 0 value point after the last recorded data is replaced by the subsequent recorded data. The average of these two points is then used to fill all 0 values within that period.

These preprocessing steps aim to clean and rectify abnormal data within the dataset, ensuring the accuracy and reliability of subsequent model training and prediction processes.

4.2 LSTM, Transformer and CNN

Initially, LSTM, Transformer, and CNN models are employed for model training and testing using historical response load data and electricity load data. These three fundamental models utilize the same dataset format, where historical electricity load data from responding users 100 days before the response day serves as input, and the response load data of responding users on the response day is set as the target.

The hyperparameters for all three basic models remain consistent, comprising a learning rate of 0.001, a learning period of 500, and the Mean Percentage Error (MPE) as the loss function. The efficacy of the test set is summarized in Table 2:

Table 2. Loss values of three basic prediction models.

Model	LSTM	Transformer	CNN
Loss	324.9591	240.5648	154.6302

In comparison to LSTM and Transformer, CNN exhibits a lower Mean Square Error (MSE) when predicting user-responsive load. Furthermore, while the loss function plots of LSTM and Transformer fail to converge effectively in this experiment, the loss function plot of CNN achieves convergence. Consequently, the experiment proceeds with the CNN model.

Given that user electricity consumption data is influenced by holidays or maintenance shutdowns, simple models may struggle to achieve accurate predictions. Therefore, this experiment introduces a Parallel sparse CNN model to enhance prediction effectiveness.

4.3 Parallel sparse CNN

Considering that users' power consumption may be influenced by holidays or shutdowns for maintenance, this paper proposes a Parallel sparse CNN model. The model consists of one CNN model for processing global data and four CNN models for processing local data in parallel. The four CNN models handling local data incorporate historical electricity load data from the 100 days preceding the target day at specific time steps. The resulting prediction outcomes are then averaged and weighted against the prediction outcomes of the CNN model processing global data to generate a weighted output.

Since the Parallel sparse CNN model integrates multiple local CNN models with a global CNN model, the predicted values from both components are calculated based on a predetermined weight ratio. Ultimately, a predicted value responsive to the load data is obtained.The configuration of this model is illustrated in Table 3:

Table 3. Parallel CNN model configuration.

Model configuration		
Global CNN	kernal_size	100
	stride	100
	linear_output_size	
Sparese local CNN	kernal_size	25
	stride	25
	linear_output_size	1
learning_rate	-	0.001
learning_epoch	-	500

The ablation experiment of weights is shown in Table 4:

Table 4. Loss values with different weights.

Sparese local CNN	Global CNN	Loss
1	9	120.7623
2	8	112.5576
3	7	99.4350
4	6	92.8640
5	5	87.9645
6	4	85.2456
7	3	93.8846
8	2	108.0312
9	1	130.6302

Based on the outcomes of our weight ablation experiments, table 4 presents the loss values corresponding to various combinations of local CNN and global CNN weights. It is evident that the smallest loss value occurs when the local CNN weight is set to 6 and the global CNN weight is set to 4. This finding suggests that this particular weight combination yields optimal model performance.

Table 5. Performance comparison of two prediction methods.

Model	RMSE	MAE	MAPE
CNN	35.27	26.01	3.35%
Parallel CNN	28.71	21.57	3.11%

Table 5 presents the prediction error values of two models on the test dataset, highlighting the superiority of the Parallel CNN model over the CNN model across all evaluation metrics: RMSE, MAE, and MAPE. Specifically, RMSE gauges the deviation between predictions and actual observations, while MAE provides a measure of the average absolute error. Additionally, MAPE offers insights into the average percentage of prediction errors relative to actual values.

Comparison between the predicted values and real values of the Parallel CNN model and the CNN model, as shown in Fig. 2.

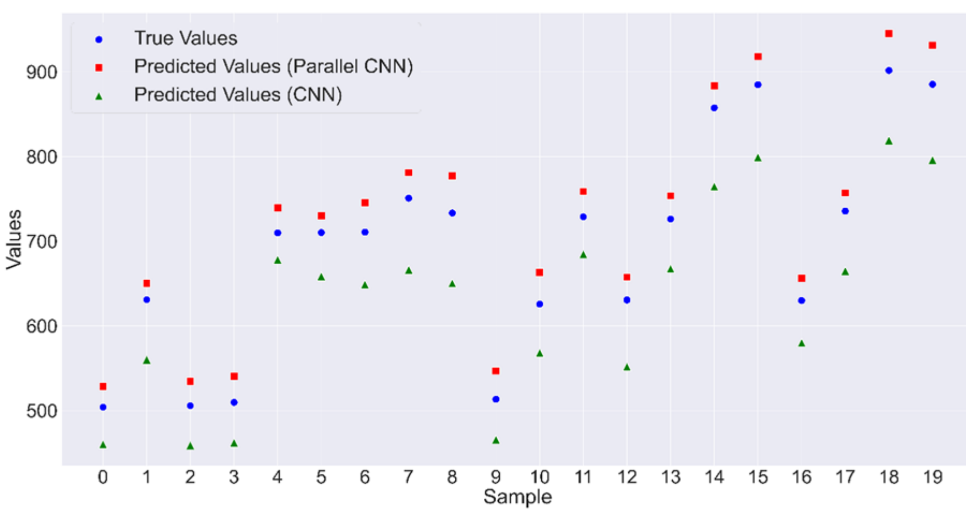

Fig. 2. CNN VS Parallel CNN.

For our comparative analysis, we selected historical responsive load data from a cohort of 20 users engaged in demand response. In Fig. 2, the triangles represent the user response load values forecasted by the CNN model, while the squares denote predictions from the Parallel CNN model. The circles illustrate the users' historical response load values. Notably, the predictions generated by the Parallel CNN model exhibit a closer proximity to the users' historical response load data compared to those made by the traditional CNN model. This indicates that the Parallel CNN model outperforms the conventional CNN model in predicting user response load values.

The effectiveness of the Parallel CNN proposed in this study is readily apparent, as it adeptly manages abnormal power consumption scenarios stemming from user holidays and line maintenance downtime. Consequently, models trained using this approach exhibit exceptional accuracy in forecasting high-voltage power consumption patterns.

4.5 EM data augmentation

Given the scarcity of existing historical response load data, the trained model's applicability is limited, lacking robustness and exhibiting relatively low prediction accuracy. To address this, this paper employs a weak supervision method to judiciously augment the data, with the augmentation process following the EM algorithm.

In E-Step, the process commences by training a basic model using the available historical response load data. This basic model is designed to capture the characteristics of the existing data comprehensively. Subsequently, the trained basic model is utilized to predict all samples, yielding the model's predictions for each sample's response load. These predictions are then used to label the data accordingly.Following the labeling process, the newly labeled data is utilized to train a new model. Each iteration of the EM algorithm is illustrated in Fig. 3.

Fig. 3. Loss iteration value.

As depicted in the figure, during the 7th, 8th, and 9th iterations, it becomes evident that the Loss value converges, signifying that the model has been trained to its optimal state. Furthermore, comparison with the model initially trained using the original response load data reveals a significant reduction in Loss value.

5 Conclusion

The model proposed in this article not only addresses the challenge of effectively training prediction models due to limited response load data but also demonstrates remarkable accuracy in predicting load behavior during user holidays or circuit maintenance periods. This innovative approach tackles the scarcity of response load data by leveraging limited information effectively, resulting in more precise predictions of user behavior.

Furthermore, the model's versatility extends to scenarios where power consumption decreases during holidays or circuit maintenance. Even in such circumstances, where abnormal power consumption data may arise due to significant reductions in high-voltage user consumption compared to regular working periods, the trained model remains capable of providing accurate load response predictions. This adaptability significantly enhances the model's practicality and reliability in real-world applications, marking substantial progress in the field of power load forecasting.

Overall, the model's ability to address data scarcity and accurately predict load behavior under varying conditions makes it a valuable asset in power management applications, offering tangible benefits to stakeholders in the power industry.

This work was supported in part by the Project of Yichang Power Supply Company, State Grid Hubei Electric Power Co., Ltd. (B715H023001N). The corresponding author is Bitao Li.

References

1. J. H. Kim, B. S. Lee, C. H. Kim. Electr. Power Syst. Res. **221**: 109415 (2023)
2. H. Hua, M. Liu, Y. Li, S. Deng, Q. Wang. Electr. Power Syst. Res. **216**: 109057 (2023)
3. S. Zhang, R. Chen, J. Cao, J. Tan. Electr. Power Syst. Res. **222**: 109507 (2023)
4. Z. Xu, V. Selvaraj, S. J INTELL MANUF. 1-24 (2024)
5. M. Tan, S. Yuan, S. Li, Y. Su, H. Li, F. He. IEEE Trans. Power Syst. **35**(4): 2937-2948 (2019)
6. A. Wan, Q. Chang, A.L.B. Khalil, J. He. ENERGY. **282**: 128274 (2023)
7. Z. Tian, W. Liu, Jiang W, C. Wu. ENERGY. 130666 (2024)
8. D. Li, Y. Tan, Y. Zhang, S. Miao, S. He. Int. J. Electr. Power Energy Syst. **146**: 108743 (2023)
9. P. Ran, K. Dong, X. Liu, J. Wang. Electr. Power Syst. Res. **214**: 108885 (2023)
10. L. Ma, L. Wang, Z. Liu. IEEE Trans. Power Syst. **37**(3): 2019-2031(2021)