

RESEARCH ARTICLE

A Spatio-temporal Graph Convolutional Approach to Real-time Load Forecasting in an Edge-enabled Distributed IoSG Energy System[†]

Qi Liu^{†1} | Li Pan^{†1} | Xuefei Cao² | Jixiang Gan¹ | Xianming Huang^{*3} | Xiaodong Liu⁴

¹School of Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China

²School of Cyber and Information Security, Xidian University, Xi'an, 710071, China

³School of Computer Science, Hunan University of Technology, Zhuzhou 412007, Hunan, China

⁴School of Computing, Edinburgh Napier University, Edinburgh, EH10 5DT, UK

Correspondence

*Xianming Huang. Email: huangxianming@hut.edu.cn

[†]Both are first authors due to their equal contribution to this work.

Abstract

As the edge nodes of the Internet of Smart Grids (IoSG), smart sockets enable all kinds of power load data to be analyzed at the edge, which create conditions for edge calculation and Real-time (RT) load forecasting. In this paper, an Edge-Cloud computing analysis energy system is proposed to collect and analyze power load data, and a combination of graph convolutional network with LSTM, called KGLSTM is used to achieve mid-long term mixed sequential mode RT forecasting. In the proposed Edge-Cloud framework, distributed intelligent sockets are regarded as edge nodes to collect, analyze and upload data to cloud services for further processing. The proposed KGLSTM network adopts a double branch structure. One branch extracts the data characteristics of mid-short term time-series data through an encoding-decoding LSTM module; the other branch extracts the data features of long term timing data through an adapted graph convolutional network (GCN). GCN is used to extract spatial correlations between different nodes. In addition, by combining a dynamic weighted loss function, the accuracy of peak forecasting is effectively improved. Finally, through various experimental indicators, this paper shows that KGLSTM and weighted KGLSTM have achieved significant performance improvement over recent methods in mid-long term time-series forecasting and peak forecasting.

KEYWORDS:

Real-time Intelligent Methods, Time-series Forecasting, Graph Convolutional Network, Long Short-Term Memory, Energy Systems

1 | INTRODUCTION

Under the background of the various power acquisition equipment and the rapid development of the IoSG, Intelligent equipment is gradually used for load data RT acquisition at various power consumption sides^{1,2,3,4,5}. The popularity of smart sockets creates favorable conditions for using edge computing when processing and analyzing a large number of power load data^{6,7}. Since the end of the epidemic, the society has resumed business and production, and the load power consumption of the whole society has increased sharply. At present, few companies embed efficient AI algorithms in the products of the power industry. Therefore, it needs more practice and efforts to apply the improved RT algorithms to energy sustainable projects. Further promoting supply side reform and strengthen demand side management, especially the construction of load RT monitoring energy systems, will provide powerful help to ensure stable operation of load detection system and improve energy utilization^{8,9}.

The power industry is one of the large-scale infrastructures that put forward the informatization of physical systems earlier, but the efficient integration of demand-side power data in terms of energy generation, distribution, management, and control, and feedback control is still in the exploratory stage. With the optimization of acquisition equipment and RT intelligent methods, edge computing is gradually used in demand-side

[†]A Spatio-temporal Graph Convolutional Approach to Real-time Load Forecasting in an Edge-enabled Distributed IoT Energy System

power data analysis. The collected load data is uploaded to the cloud service in the form of edge nodes, and is analyzed and stored through the cloud server^{10,11,12}. Among them, RT load forecasting is an essential part of load monitoring. also, RT load forecasting is an important basis for power security dispatching, and it provides a reliable guarantee for the stable, reliable and economic operation of the power system¹³. The forecast objects include user load electricity consumption, annual and monthly electricity consumption and daily electricity load change process. In addition, the power data after statistical analysis can also provide a basis for the intelligent configuration and planning of the State Grid, and ensure the safety and reliability of the power grid operation^{14,15}. In the process of load data collection, due to the influence of communication quality, sensor performance, network delay and other factors, there are sometimes abnormal jump values in the load data. This kind of data that deviates from the normal value generated by the detection equipment or the transmission process is called "wild value"¹⁶. The existence of outliers seriously affects the feature extraction of data. Therefore, eliminating outliers is an important preliminary step for analyzing load data. Kalman filter has a good data cleaning effect, its filtering effect is good, and its RT performance has been confirmed¹⁷.

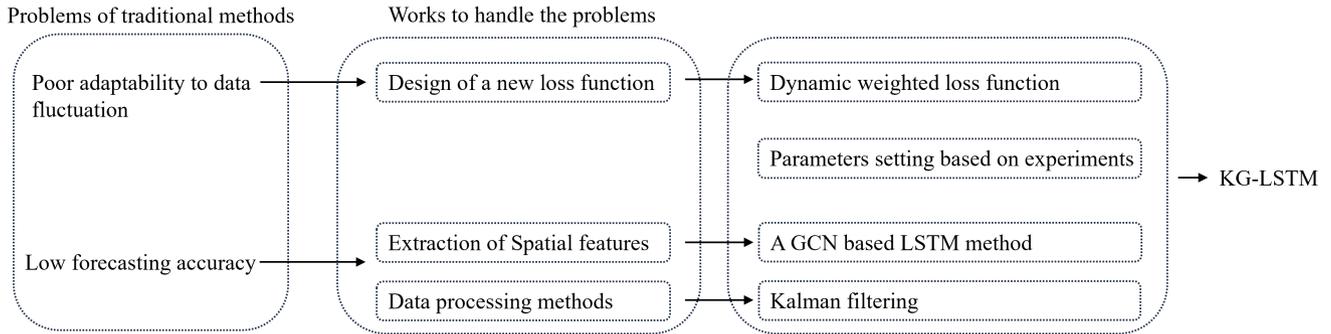


Figure 1 Motivational Diagram.

Traditional RT load forecasting method has many disadvantages, such as simple forecasting method, low forecasting accuracy and poor ability to process real-time data, which cannot meet demands of the current power grid energy system. At the same time, traditional RT load forecasting method can only deal with forecasting situation with small fluctuation of load data. For short term RT load data, effect of traditional RT forecasting method is not good¹⁸. However, with the improvement and development of various intelligent algorithms, deep learning algorithms have been widely used in RT forecasting¹⁹. As shown in figure. 1, this paper combines GCN and LSTM to forecast load energy consumption data. From the perspectives of time and space, analyzing characteristics of load energy consumption data, and making mid-short term RT load forecasting.

2 | RELATED WORK

2.1 | Big Data Quality Improvement

Power load has characteristics of non-linearity and multi time series scale, and it also shows certain regular fluctuation under the influence of meteorological conditions and other factors²⁰. When considering influencing factors, two points should be paid attention to: First, Missing factors will lead to one-sided feature extraction and affect forecasting accuracy. Second, different load data have different load characteristics, and the main influencing factors should be screened according to various load characteristics. However, both influencing factors and load data all have missing values, abnormal values, noise and other abnormal values²¹. However, Data cleaning has received huge attention in the field of big data algorithms. Data cleaning is mainly to solve the inconsistency problem by filling in missing values, smooth noise data, identifying or deleting outliers and other methods²². There are three commonly used data cleaning methods, namely binning method, clustering method, and regression method²³. The Kalman filter method in the regression method used in this paper. Yang et al. in 2009 used Kalman filter to process the forecasted gyro signal, which effectively improved the forecasting accuracy²⁴. Behrouz et al. in 2013 used Kalman filter to optimize the overall accuracy of the output data of the ANN algorithm, and compared Kalman filter with other filtering methods²⁵. Various methods have proved that Kalman filter has an effective effect on data cleaning in big data algorithms.

2.2 | Development of RT Load Forecasting

RT load forecasting algorithm can be divided into the following categories: conventional statistical RT forecasting method, RT load forecasting method based on machine learning and RT load forecasting method based on artificial intelligence²⁶.

Conventional statistical RT forecasting methods include regression analysis method²⁷, multiple linear regression method²⁸, and gray coefficient forecasting method²⁹. The regression analysis method uses the principle of data statistics to realize the forecasting of the data by establishing the variables³⁰. Multiple linear regression is a method of using multiple variables to achieve linear regression, and is An improved single variable linear regression algorithm. Moreover, the accuracy of multiple linear regression in RT load forecasting is much higher than that of a single linear regression method^{31,32}. The regression model has a good forecasting effect for stationary sequences and high accuracy, and the model itself has a relatively simple structure. However, for nonlinear and complex sequence data, the regression model has a general effect and weak self-learning ability³³. Grey forecasting is the use of sequence operators to calculate the system evolution law of the original data, establish a grey system model, and make quantitative forecasting for the future state of the system. It was first proposed by Professor Deng. This model has low data requirements, and has a good forecastive effect on data with small amounts of data and many uncertain problems³⁴.

RT forecasting methods based on machine learning include SVM; Xgboost et al. SVM is a useful machine learning model for text classification proposed by vapnic et al in 1964. It is a generalized linear classifier based on VC dimension theory and risk minimization³⁵. Xing Yan et al constructed a new machine learning model by combining SVM and ARMAX to forecast the PJM power grid data. Compared with SVM and ARMAX, SVM-ARMAX improves forecasting accuracy by calculating linear modules, and optimizes shortcomings of SVM in RT time series forecasting³⁶. De Yue men et al in 2011 proposed an improved support vector machine method: LS-SVM and made forecasting using the data set of local history for 5 years³⁷. This method takes into account the trend component and the period component by introducing the SVM. Which improved accuracy of time series forecasting.

Xgboost is an improved gradient lifting decision tree machine learning algorithm based on GBRT proposed by Chen and guestrin³⁸. At present, Xgboost has been widely used in various fields³⁹. At the same time, scholars in various fields have adopted combinatorial optimization methods to ensure the efficiency and accuracy of RT data forecasting models. Deng et al used bagging and Xgboost to analyze the relationship between short term load data and weather influencing factors. At the same time, weighted similarity selection is substituted to improve RT load forecasting accuracy of the model⁴⁰. Li et al constructed a RT sequence forecasting model based on Xgboost to make accurate RT forecasting for the instability of crustal plate movement⁴¹.

Compared with conventional statistical RT forecasting methods, advantage of RT forecasting methods based on machine learning is that it can deal with both linear and nonlinear time series. What's more, RT forecasting methods using machine learning can effectively avoid the situation of data overfitting and the problem of local minimum caused by too large sample data⁴². However, when using RT forecasting methods using machine learning to forecast the time series data, complexity of the method could be high, and the time response is long for the second level RT response.

With the update of technology, artificial intelligence methods had been continuously introduced into RT load forecasting, and forecasting methods based on artificial intelligence forecasting have gradually received attention. The forecasting method based on artificial intelligence methods can be divided into single neural forecasting network and combined neural forecasting network⁴³.

2.3 | RT Load Forecasting Methods Using Single Neural Network

Commonly used single neural networks are CNN, LSTM, etc. CNN network was proposed by LeCun, Y et al in 1989. The local connection and weight sharing method adopted by this model greatly improved the optimization degree of the network while reducing the risk of overfitting⁴⁴. Maryam used CNN to extract the nonlinear relationship between power load data and relevant temperature. after model learned the hidden nonlinear load temperature characteristics, the accuracy of RT residential load forecasting could be improved⁴⁵. LSTM network is an improvement of RNN model that can learn long term dependency information. Of course, LSTMs and baseline RNNs are not particularly structurally different, but they use different functions to compute "hidden" states^{46,47,48}. At the same time, in 2005, Alex Graves et al proposed an optimized LSTM network, namely the bidirectional long term and short term memory network (vailla LSTM). Finally, this network developed into the neural network with the best use effect at present⁴⁹. This network improves the advantages of the reserved RNN network, and also optimizes the disadvantages of RNN for long time series processing, which can forecast the real-time data effectively.⁵⁰ Hossain applied LSTM to RT forecasting of photovoltaic power generation, and combined the K-means algorithm to convert historical data into dynamic phenolic characteristics, so as to improve the measurement accuracy⁵¹.

Compared with RT forecasting methods based on machine learning. The RT forecasting method based on artificial intelligence has a significant breakthrough in both forecasting accuracy and response time. However, there are many kinds of RT forecasting methods based on artificial intelligence, which have their own advantages, but they cannot meet all the actual conditions. Therefore, scholars often adopt the hybrid model with artificial intelligence algorithm as the core to make more accurate forecasting of various time series.

2.4 | RT Load Forecasting Method Using Combined Neural Network

Considering that a single neural network will have various defects. Scholars try to combine various types of neural networks to learn from each other's strengths to achieve better forecasting results. The combined neural network has now become the mainstream in load forecasting. Yang et al aiming at the problem of low accuracy of RT precipitation forecasting caused by inaccuracy of radar echo motion trend and intensity. The self-attention mechanism and LSTM are combined to form a self-attention integrated recursive unit (ST-LSTM) to capture temporal and spatial characteristics of radar echoes and improved forecasting accuracy⁵². Guokun Lai et al. proposed an efficient deep learning framework: the Long Short Term Time Series Network (LSTNet). Using CNNs and RNNs to make long- and short term forecasting on data. And accurately forecast the load data of the next hour on the load power data set from 2012 to 2014⁵³. Jihua Ye et al. proposed an Attention Based Spatio-Temporal Graph Convolutional Network considering External Factors (ABSTGCN-EF) for multi-step traffic flow prediction⁵⁴. Yuan Gao et al. proposed a combined spatial-temporal Energy Harvesting (EH) and relay selection scheme, which improved the Energy Efficiency (EE) and spectrum utilization of Cognitive Wireless Powered Networks (CWPNs)⁵⁵. Shun-Yao Shih et al. proposed a self-attention mechanism neural network (TPA-LSTM) to solve the multivariate time series forecasting problem. This model incorporates attention mechanism in recurrent neural network (RNN), experiments on the typical MTS dataset and the polyphonic music dataset and forecasts the results⁵⁶.

With the convenient acquisition of more and more load data and the mature application of various network models, the application of deep learning models for RT load forecasting methods has become the most active research direction in this field.

3 | METHODOLOGY

3.1 | Data Cleaning Method Enhanced by Kalman Filter

The Kalman filtering method is a method that uses noisy original data sets to forecast the next data. Kalman filter is a linear random state space composed of state estimation and observation estimation equation. The state estimation equation is used to forecast the next data based on the previous, and the observation estimation equation is used to transform the data to a estimation vector. The state estimation equation part and the observation estimation equation part are composed of equations (1) and (2).

$$\theta_k = A\theta_{k-1} + B\mu_{k-1} \quad (1)$$

$$z_k = C\theta_k + v_k \quad (2)$$

The best state of the filter is evaluated using a regression algorithm based on a linear unbiased least mean square estimation criterion, which results in the best estimated state of the filter. The Kalman filter model consists of two states, forecasting and update.

- forecasted State:

The forecasted state is to forecast the current state according to the best forecast value of the previous state:

$$\theta_k = A\theta_{k-1} + B\mu_{k-1} \quad (3)$$

At the same time, the forecasted state begins to update the forecasted covariance:

$$P_k = AP_{k-1}A^T + Q \quad (4)$$

- Update State:

The update state first calculates the Kalman gain of the current state:

$$K_g = P_k C^T / (C P_k C^T + R) \quad (5)$$

Then modify the optimal forecasting state of the system in the k state θ'_k :

$$\theta'_k = \theta_k + K_g (z_k - C\theta_k) \quad (6)$$

Finally update the covariance p'_k of the optimal forecasting state of the system in the k state:

$$p'_k = (I - K_g C) P_k \quad (7)$$

I is the identity matrix, and for the case of a single model and single measurement, $I = 1$. For this reason, the Kalman filter algorithm continuously calculates the covariance of the optimal forecasting state and the updated state according to the system process. Through this autoregressive method, the optimal value of the system is obtained, so as to achieve the filtering effect, reducing data error of the original data, and improving the effect of later forecasting. .

3.2 | Dynamic Weighted Loss Function

Considering the uneven electricity consumption of household loads over different time periods. In this paper, a weighted MSE loss function is used to optimise this type of problem. Specifically, the corresponding weight size weight (X_i) is assigned to each load according to its intensity of electricity consumption. This is shown in formula (8). Region num: $[\text{Max}(X_i) - \text{Min}(X_i)] / 5$.

$$\text{weight}(X_i) \begin{cases} 1, & X_i < \text{Region num} \\ 2, & \text{Region num} < X_i < 2 * \text{Region num} \\ 5, & 2 * \text{Region num} < X_i < 3 * \text{Region num} \\ 7, & 3 * \text{Region num} < X_i < 4 * \text{Region num} \\ 10, & 4 * \text{Region num} < X_i \end{cases} \quad (8)$$

After optimization, the weighted loss function can be defined as: $\text{mse}_{\text{weight}} = \frac{1}{n} \sum_1^n [\text{weight}(X_i) * (X_i^P - X_i^t)^2]$

3.3 | Codec Forecasting Method Using Spatio-temporal Characteristics

The proposed codec forecasting model based on Kalman filter and GCN (KGLSTM) can be shown in Figure. 2. This model is a two-branch combinatorial forecasting network composed of LSTM decoder-encoder and GCN. The network structure consists of a data cleaning module and a feature extraction module. The input is the power load data $X_{i-m+1}^T, \dots, X_i^T$ for t time period of the nth room, first through the data cleaning module, filter out invalid data and clean the data to get the power load data $X_{i-m+1}^{T_0}, \dots, X_i^{T_0}$. Then the cleaned time feature information and spatial feature information are extracted through the feature extraction module. Finally, the final load forecasting data X_i^P is obtained through the feature fusion algorithm in the forecasting module.

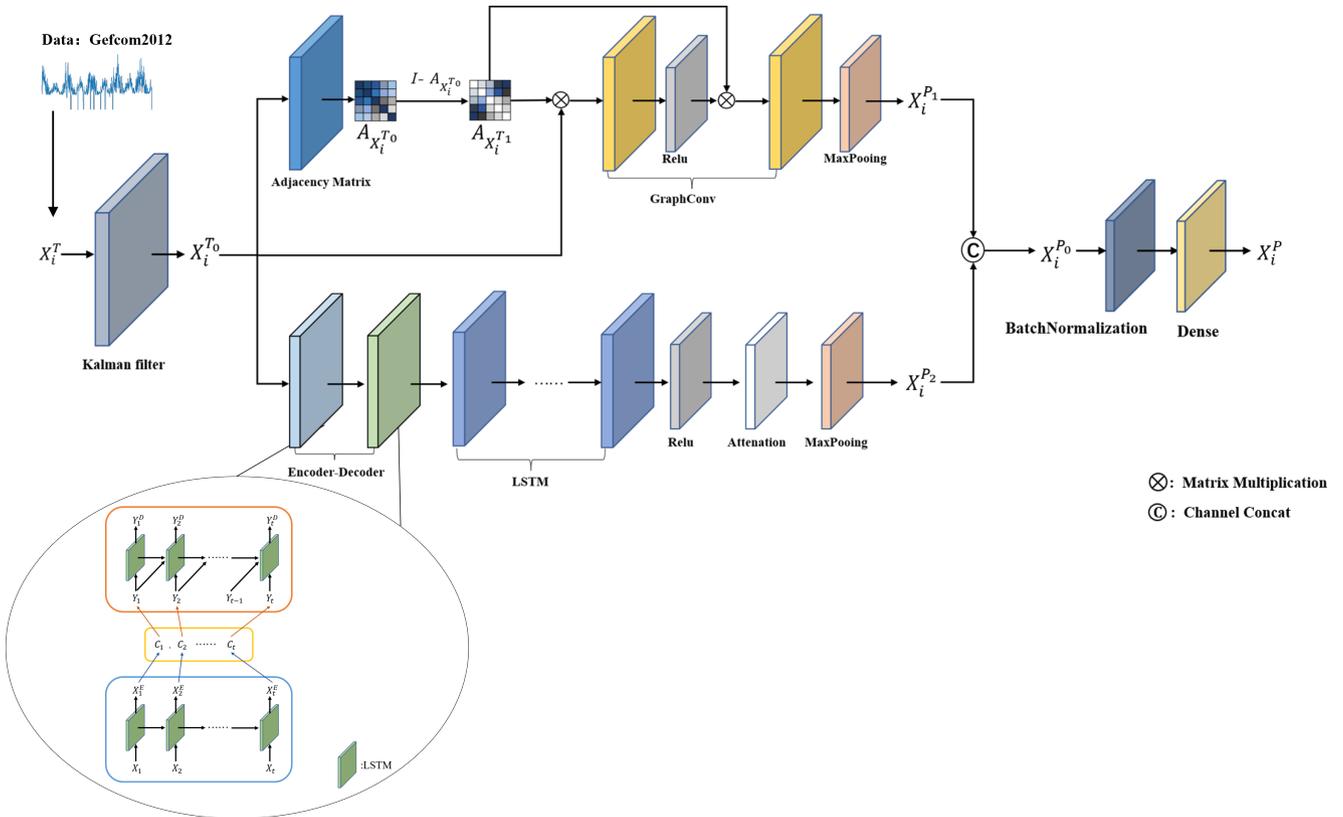


Figure 2 KGLSTM network structure diagram.

3.3.1 | Data Cleaning Module

The data cleaning module is composed of data division and Kalman filtering. The data input module divided the input data set into three categories: the first 1/2 is the training data, the second 1/4 is the test data, and the last 1/4 is the verification data. Then input the segmented data into the Kalman filter module for data cleaning to filter out the wrong data, filled in the missing data, and smoothed the data curve, so as to achieve the purpose of improving the forecasting accuracy.

3.3.2 | Feature Extraction Module

The feature extraction module including two modules. The first module is the GCN feature extraction module, and the other is the decoding time sequence feature information analysis and extraction module of the LSTM.

- The information extraction module of GCN

GCN is a deep learning network that deals with graph structure. In a graph structure, the node characteristics of each graph node are composed of the characteristics of the node itself and the characteristics of neighboring nodes around it. The goal of graph neural is to learn the state embedding of each node. Among them, the GCN, as an extension of the convolutional neural network, applies the convolution operation to the graph structure to realize the feature extraction of the graph node information. At present, there are roughly two types of graph convolution methods, one is based on Graph convolution in the spatial domain; the other is spectral-based graph convolution⁵⁷. In this paper, spectral-based graph convolution is used to extract relevant sequence feature information. The main idea of spectrum-based graph convolution is to use Fourier transform to convert the input graph signal data into spectral data, and finally perform the convolution operation. Among them, The Laplace matrix L calculated in the spectral convolution process is represented as follows:

$$L = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} = I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \quad (9)$$

In formula (9), D represents the degree matrix in the process of operation, defined as $D_{ii} = \sum_i A_{ij} \cdot I_N$, I_N is a unit matrix with a size of $N * N$. Then, decompose and extract the eigenvalues of L to obtain:

$$L = UAU^T \quad (10)$$

In formula (10), A is a diagonal matrix composed of the eigenvalues of L . Among them, $U = \{u_1, u_2, \dots, u_n\}$ is composed of the eigenvectors of L , corresponding to a set of orthogonal bases in the R^N space.

The transfer formula between spectral domain convolution operations used in this paper is:

$$X^{l+1} = GCN(A, X) = \sigma \left(\tilde{D}^{\frac{1}{2}} \tilde{A} \tilde{D}^{\frac{1}{2}} X^l W^l \right) \quad (11)$$

In formula (11), W^l represents the learning parameter and σ is the activation function. $\tilde{A} = A + I_N$ means the method of performing self-circulation operation on the adjacency matrix A itself, by adding an identity matrix, the central node itself is also added to the convolution operation. $\tilde{D} = \sum_i \tilde{A}_{ij}$; $X^l \in R^N$. The structure diagram of the GCN based on the spectrum can be shown in Figure. 3.

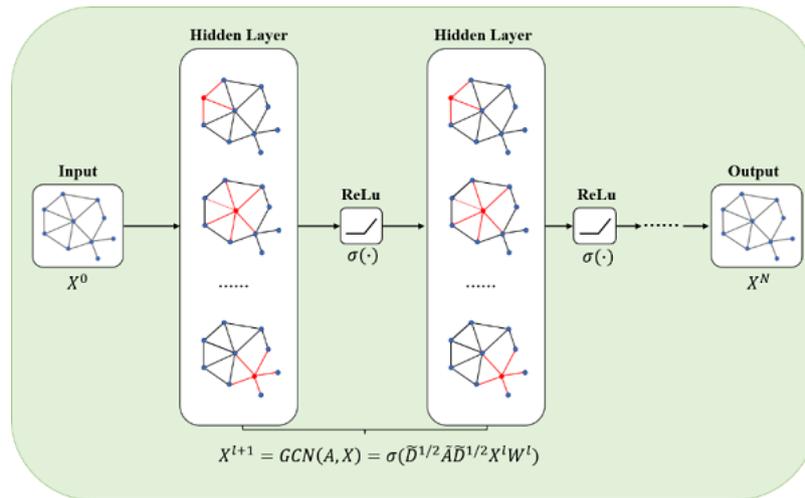


Figure 3 Spectrum-based GCN Structure Diagram.

Encoder-Decoder is a sequence-to-sequence model that can be combined with commonly used forecast time series models such as LSTM and RNN. In recent years, the network model combining encoder decoder and time series model has been applied to all kinds of long and short time series forecasting. For example, Yao Qin et al⁵⁸. combined Encoder-Decoder with LSTM in DA-RNN to forecast indoor temperature. Mina Razghandi and others⁵⁹ combined Encoder-Decoder and LSTM to establish a sequence-to-sequence forecasting model. Used to forecast the load power consumption of various household appliances.

This paper combines Encoder-Decoder and LSTM to extract the short term load power consumption characteristics of each room. Taking LSTM as the basic unit, the encoder and the decoder are connected to each other through a fully connected layer, and finally the extracted feature information is restored to the dimension information output consistent with the input dimension. The model structure can be shown in Figure. 4.

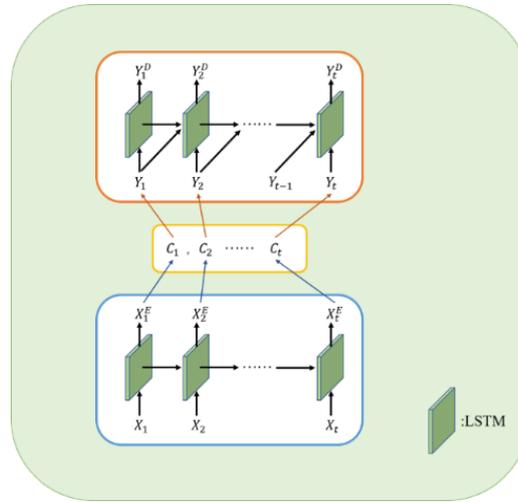


Figure 4 Encoder-Decoder Structure Diagram Combined with LSTM.

Although the LSTM encoder-decoder is based on the overall feature extraction effect of time series data, but for some special value (maximum value, minimum value) features can't be accurately extracted, considering that the span of this data set is long for 2-3 years, to extract the features in the data set accurately, This model adds the attention machine model to the multi-layer LSTM module to extract the salient features in the time series data.

The calculation steps of the attention algorithm are as follows: First, calculating the attention distribution rule of input sequence $Input_X : [x_1, x_2, \dots, x_n]$. The second is to obtain the relevant weighted average information by calculating the input sequence $Input_X : [x_1, x_2, \dots, x_n]$ according to the obtained attention distribution, by giving the input sequence $Input_X : [x_1, x_2, \dots, x_n]$ gives different weights, and reflects the importance of various information in the sequence through the relationship between the weights.

In order to find information related to a certain feature in the sequence $Input_X : [x_1, x_2, \dots, x_n]$, first introduce a representation information related to the task, denoted as Q, and pass the scoring function to calculate the correlation between each input quantity and Q in the input sequence $Input_X : [x_1, x_2, \dots, x_n]$. The common scoring functions are dot product operations, scaled dot product operations, and bilinear operations, respectively, as shown in formula (18), formula (19), and formula (20):

$$\text{Similarity}(Q, X_i) = Q \cdot X_i = X_i^T Q \tag{12}$$

$$\text{Similarity}(Q, X_i) = \frac{X_i^T Q}{\sqrt{d}} \tag{13}$$

$$\text{Similarity}(Q, X_i) = X_i^T W Q \tag{14}$$

Depending on the specific generation method, the numerical value range is also different. In step 2, a calculation method similar to *Softmax* is introduced to perform numerical conversion on the score calculated in step 1. Generally, the following formula is used to calculate:

$$a_i = \text{Softmax}(\text{Similarity}(Q, X_i)_i) = \frac{e^{\text{Similarity}(Q, X_i)_i}}{\sum_{j=1}^{L_x} \text{Similarity}(Q, X_j)_j} \tag{15}$$

Finally, the obtained attention distribution a_i and the input sequence $Input_X : [x_1, x_2, \dots, x_n]$ are summarized to obtain:

$$\text{attention}(X, Q) = \sum_{n=1}^N a_n \cdot x_n \quad (16)$$

The algorithm structure of attention mechanism can be shown in Figure. 5.

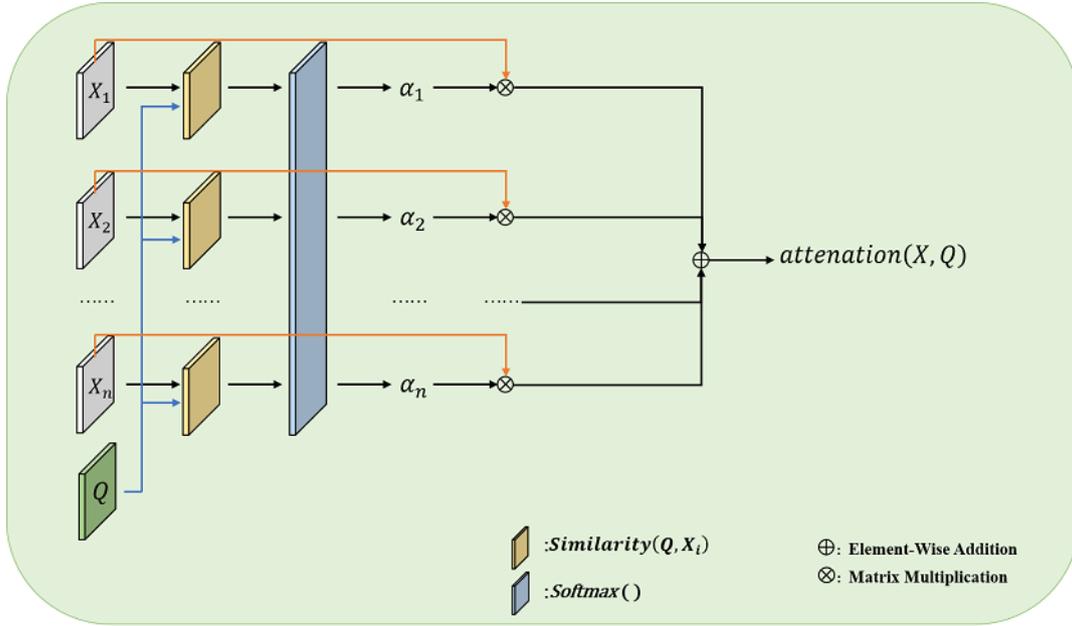


Figure 5 The algorithm structure of the attention mechanism.

The initial input X_i^T , after Kalman filtering, $X_i^{T_0}$ is obtained, and the cleaned $X_i^{T_0}$ get $X_i^{P_1}$ and $X_i^{P_2}$ respectively through the space load feature extraction layer of GCN and time load feature extraction layer of encoding and decoding LSTM, and linearly stitches $X_i^{P_1}$ and $X_i^{P_2}$ to obtain the final integrated feature data $X_i^{P_0}$. Finally, according to the load feature $X_i^{P_0}$ obtained by the spatial load feature analysis layer of the GCN and the time load feature processing layer of the encoding and decoding LSTM, the input load data at the next moment is forecasted to achieve the purpose of improving the forecasting effect.

3.4 | Edge-cloud Computing Test Bed

Based on IoSG, In the Edge-Cloud computing energy system, each unit is installed with corresponding specifications of smart sockets. These smart sockets are installed in each room of each unit and can accurately record the load power consumption of each room. the Edge-Cloud computing energy system uses each building as an edge node and each room as a bottom node to collect load power data in each building. At the same time, The collected load power data is transmitted to the cloud server for load monitoring and analysis⁶⁰. Finally, the collected load power consumption data is stored in the mirror backup server for retention on the one hand. On the other hand, the forecasting data is obtained through the analysis of the cloud server, and then uploaded to the data storage server through the cloud service. At the same time, for some warning data exceeding the threshold value, the cloud server will directly send the warning signal to the smart socket to avoid risks⁶¹. The cloud server consists of data preprocessing and data forecasting modules. The data forecasting module obtains the corresponding forecasting data from the preprocessed load data through the deep learning algorithm, and feeds it back to the corresponding users through the Web application interface. The frame diagram of the Edge-Cloud computing analysis platform can be shown in Figure. 6.

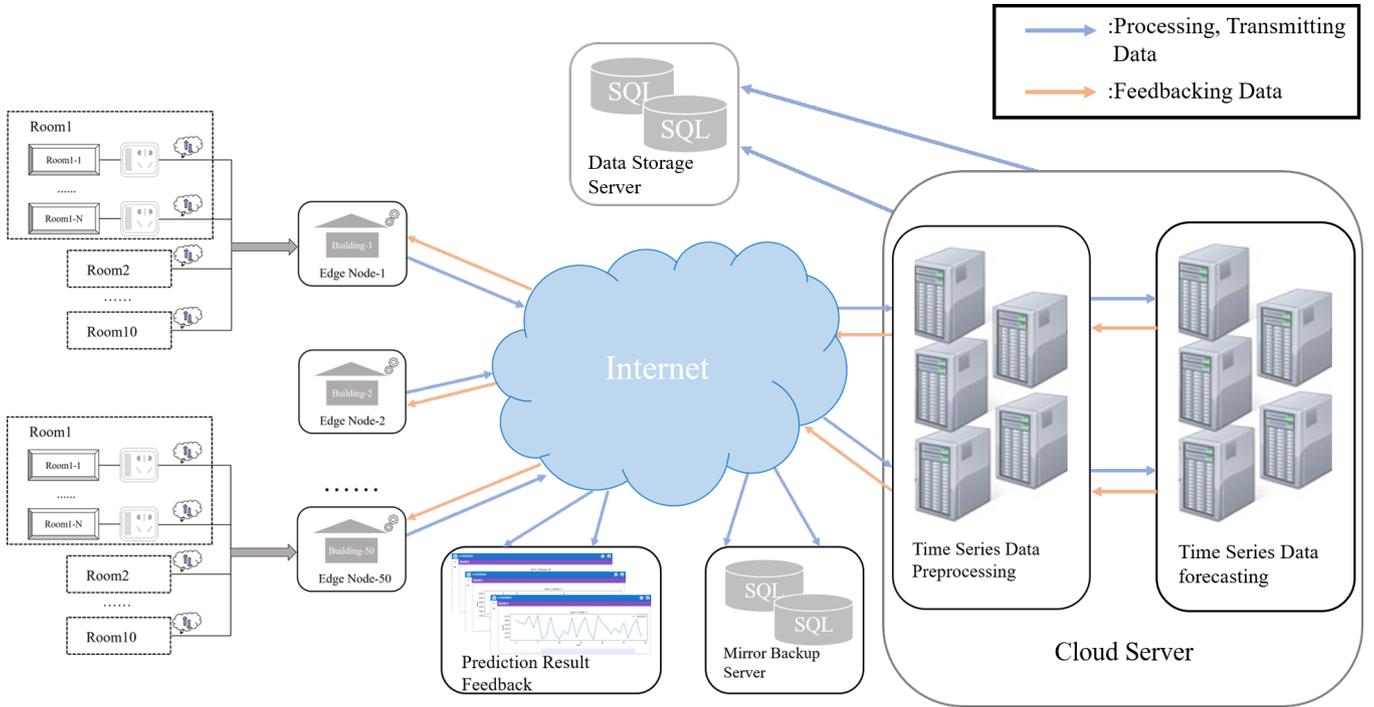


Figure 6 Framework diagram of Edge-Cloud computing analysis platform.

4 | EXPERIMENT EVALUATION

4.1 | Graph Generation Method for Spatial Correlation

For the feature extraction of load power data, researchers will consider the various influencing factors on the load power. The mainstream influencing factors are: season, temperature, number of users, and house location^{62,63}. Among these consideration factors, Few people consider the relationship between houses and houses. This model starts from the relationship between the house A and the house B, and generates the corresponding topological structure network for the relationship between the house A and the house B. It proves that the relationship between the house A and the house B is also one of the influencing factors. Feature extraction, and finally extract the load power characteristics of each house. First, calculate the pearson correlation coefficient between the house and the house in the input load power time series data, that is, the formula(23):

$$R = \frac{\sum_{i=n}^n (P_i - \bar{P}) (Z_i - \bar{Z})}{\sqrt{\sum_{i=n}^n (P_i - \bar{P})^2} \sqrt{\sum_{i=n}^n (Z_i - \bar{Z})^2}} \quad (17)$$

Among them, P and Z are the variables that need to be brought into the calculation. The final result is a symmetric matrix K with dimension $N * N$. This paper takes 10 houses with building number 1 and room numbers from 1 to 10 as examples, and draws the pearson correlation coefficient diagram between each house and the house, which can be shown in Figure. 7.

Considering that the obtained pearson correlation coefficient matrix K contains spontaneous node information, the pearson correlation coefficient matrix K is transformed into the adjacency matrix K' , by formula (24), the correlation formula is as follows, where I_0 is the identity matrix.

$$K' = I_0 - |K| \quad (18)$$

After obtaining the adjacency matrix K' between the house and the house, K' is normalized by Laplacian to perform eigen-decomposition. The specific process is shown in formula (25-27).

$$\tilde{K} = K' + I_N \quad (19)$$

$$D_{ii} = \sum_j \tilde{K}_{ij} \quad (20)$$

$$\hat{K} = D^{\frac{1}{2}} \tilde{K} D^{\frac{1}{2}} \quad (21)$$

The decomposed feature information is brought into formula (28) for convolution calculation:

$$A^{l+1} = GCN(K, A) = \sigma \left(\tilde{D}^{\frac{1}{2}} \tilde{K} \tilde{D}^{\frac{1}{2}} A^l W^l \right) \quad (22)$$

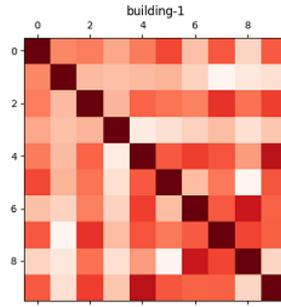


Figure 7 Pearson correlation coefficient diagram between houses 1-1 to 1-10.

Finally, from No. 1 to No. 10, the obtained graph structure information of the 10 rooms can be brought into the GCN network for feature extraction from the perspective of spatial correlation.

5 | DATASETS

The data set chosen in this experiment is the daily power consumption data of 500 houses in a city from January 1, 2013 to December 31, 2017. The data set contains daily power consumption data in some communities, with 1 to 50 buildings per building, each of which contains 10 houses. The forecasting objects are the daily power consumption data of building number 1 and house number. The training set selects the first 3/5 data in the load data set, the test set selects the next 1/5 data, and the verification set selects the last 1/5 data.

At the same time, in order to confirm the generality of this paper proposed method and the reliability of the data set, the data set of the 2012 global energy forecasting competition is also used in this experiment(<https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting>),The training set selects the first 3/5 data in the load data set, the test set selects the next 1/5 data, and the verification set selects the last 1/5 data.

6 | EVALUATION INDICATORS

To verify the validity of the proposed model, this paper selects 6 types of indicators to quantify the relevant experimental results. On the one hand, accuracy of the model is measured by using the R^2 ($R^2 - Score$); $RMSE$; MAE and MSE . On the other hand, considering that peak forecasting is very important to RT load forecasting, this paper proposes a new type of peak forecasting standard: ($Peak - Score$) to measure the accuracy of peak forecasting by various models.

- Peak forecasting: $Peak - Score$

Because TS index cannot judge the validity of the forecasted value beyond the real data. Therefore, a new peak index $Peak - Score$ has been proposed, which combines the threat score (TS) and the coefficient of determination (R^2). the specific formula can be seen in formula (29). Among them, the setting range of θ is between 0.4-0.6.

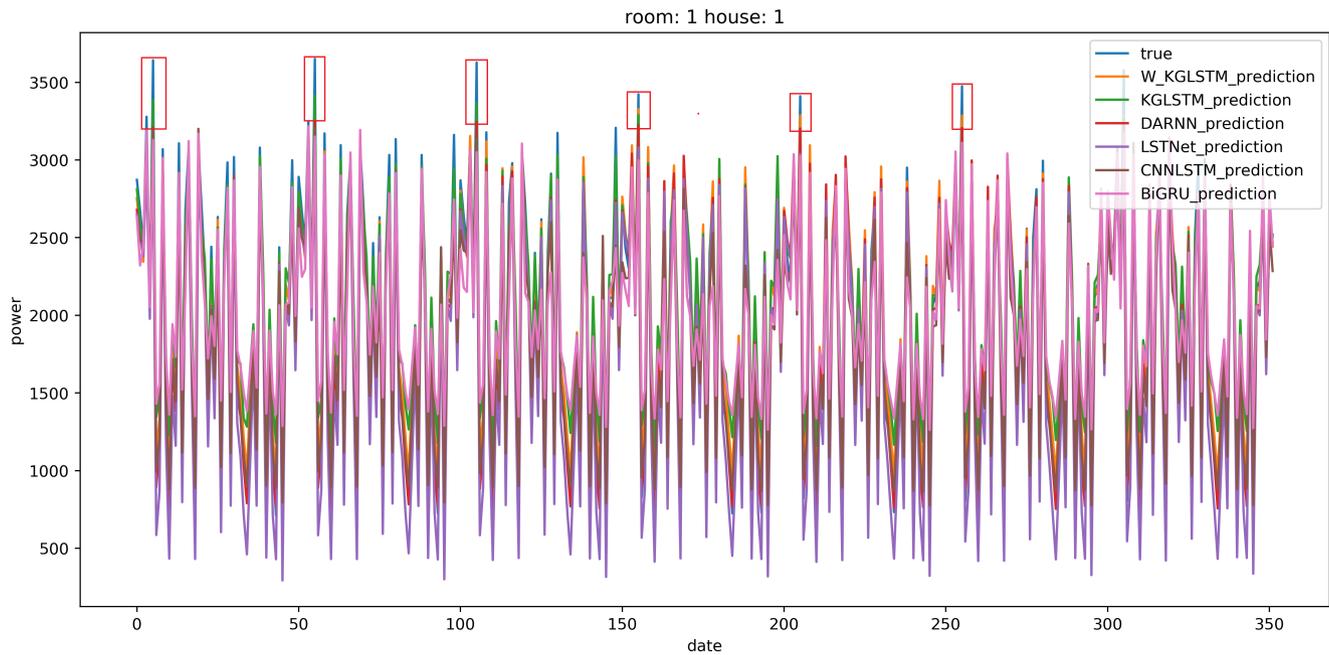
$$Peak - Score = \theta * R^2 + (1 - \theta) * TS \quad (23)$$

6.1 | Comparison Model

In order to verify the reliability of the model which proposed in this paper, four kinds of optimization models for LSTM: BiGRU, CNN-LSTM, LSTMNet and DARNN, are selected, and compared with the KGLSTM and weighted KGLSTM (W-KGLSTM) proposed in this paper. The forecasting results of various models for the two data sets can be shown in Figure. 8 and Figure. 9. The first type of data set selects the five-year power load data of a city with building number 1 and house number 1, and selects the data of the fourth year to forecast the data of the fifth year. The second type of data set selects the four-year power load data of a city with building number 1 and house number 3, and selects the data of the third year to forecast the data of the fourth year. The number of units in the LSTM layer of the proposed W-KGLSTM will affect the accuracy of prediction. By

Table 1 MAE loss of W-KGLSTM based on various settings of LSTM units

Number	32	64	96	128	160
MAE	0.10591	0.13017	0.08675	0.08297	0.09360

**Figure 8** Six types of RT load forecasting results based on a city's five-year data.

conducting experiments with different number settings on the GEFCom2012 dataset, it was determined based on the MAE evaluation index that the optimal prediction effect was achieved when the number of units was set to 128. Table 1 shows the results.

As shown in Figure. 8, the comparison effect between various models and KGLSTM is shown. From the prominent part of the red box, it is forecasted that the peak power consumption is higher than 3000 Watt hours. The forecasting results of BiGRU, CNN-LSTM, DARNN and LSTNet are not as good as those of KGLSTM. When the power consumption is lower than 1000 watts, the effect of KGLSTM is slightly lower than that of DARNN and CNN-LSTM. There is obvious error in the forecasting effect of LSTNet, and the forecasting effect of BiGRU is the worst. On the whole, the forecasting effect of KGLSTM has obvious advantages over that of BiGRU, CNN-LSTM, DARNN and LSTNet. Meanwhile, while retaining the advantages of KGLSTM, W-KGLSTM can better forecast the peak power consumption of users by weighting the power consumption of each interval. After weighted processing, the forecasting effect of W-KGLSTM is better than KGLSTM for data with power consumption lower than 1000 watt hours and data with power consumption higher than 3500 watt hours. It is added that for the forecasting data of the first 150 days in the red box, the performance of forecasting in W-KGLSTM is better than KGLSTM. For the 150 days forecasting data in the red box, W-KGLSTM outperforms KGLSTM in forecasting. The reason is that KGLSTM combines the advantages of GCN and LSTM coder-decoder. Considering the gradient vanishing problem of LSTM for mid-long term and long term load forecasting, resulting in the decline of forecasting accuracy.

As shown in Figure. 8, the comparison effect between various models and KGLSTM is shown. Throughout the load forecasting curves of various models, the RT load forecasting of KGLSTM and W-KGLSTM is more accurate than the other four types of forecasting curves, and the accuracy of various wave peaks is also higher. Although DARNN can show high forecasting results for the peak value, the overall forecasting accuracy is low and there is a high deviation from the real data. LSTNet, BiGRU and CNN-LSTM perform generally in both peak forecasting and precision forecasting. Compared with the first data set, the forecasting effect on GEFCom2012 data set, W-KGLSTM is better than KGLSTM in all time periods because comparing with the load value of GEFCom2012.the peak value of the first type of data set is much smaller than that of GEFCom2012. Although the forecasting effect of LSTM is excellent in the short term, the forecasting effect is lower than the weighted forecasting model for data with high load value. Finally, the evaluation indicators in Chapter 4.2 are introduced to show the performance differences of the models, as shown in Table 2.

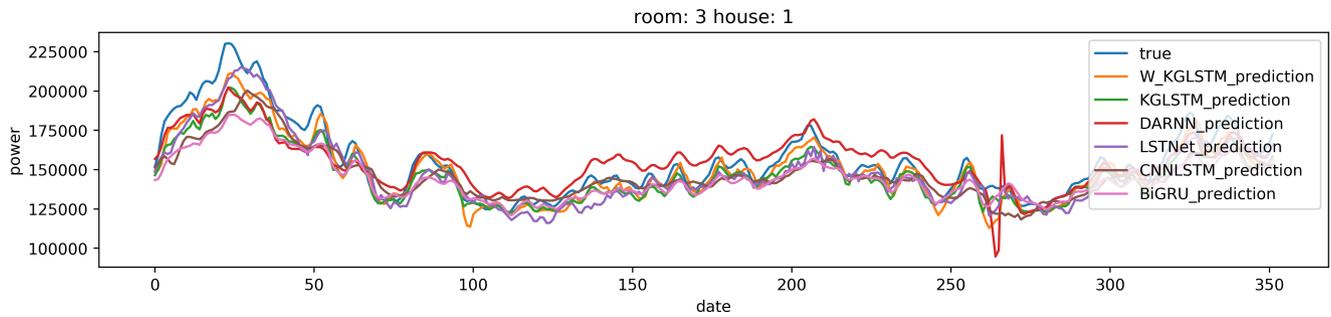


Figure 9 Six types of RT load forecasting results based on gefcom2012 data.

Table 2 Results of various evaluation indicators based on the two data sets

Model	Five year dataset of a city						GEFCom2012 dataset					
	(Peak - Score)	(R2 - Score)	(TS)	(RMSE)	(MAE)	(MSE)	(Peak - Score)	(R2 - Score)	(TS)	(RMSE)	(MAE)	(MSE)
BiGRU ⁶⁴	0.45283	0.75707	0.25	0.10494	0.078029	0.0110	0.27719	0.69297	0.0	0.11886	0.10334	0.01413
CNN-LSTM ⁶⁵	0.48536	0.87367	0.22727	0.07556	0.05847	0.0057	0.37402	0.75505	0.12	0.10558	0.0898	0.01114
LSTNet ⁵³	0.48370	0.91453	0.16666	0.04289	0.03245	0.0018	0.49106	0.91264	0.21	0.06340	0.04873	0.00401
DA-RNN ⁵⁸	0.56581	0.95925	0.33333	0.06185	0.04590	0.0038	0.46538	0.43346	0.32	0.21808	0.20075	0.04755
W-KGLSTM	0.75371	0.96763	0.61111	0.03847	0.02996	0.00148	0.64340	0.82850	0.52	0.08883	0.07824	0.00789
KGLSTM	0.67833	0.98994	0.47058	0.02136	0.01723	0.00045	0.50731	0.96827	0.2	0.03820	0.03057	0.00145

Table 2 shows that for the first type of data set (five-year data set of a city), the indicators of $R2 - score$, MAE , MSE and $RMSE$ of KGLSTM are higher than those of BiGRU, CNN-LSTM, LSTNet and DARNN, and the TS and $Peak - Score$ of KGLSTM are also greater than those of BiGRU, LSTNet, CNN-LSTM and DARNN. At the same time, in peak forecasting, the weighted loss KGLSTM (W-KGLSTM) has shown obvious advantages. Compared with KGLSTM, TS number improved by 14.052% and $Peak - Score$ number improved by 7.538%. However, indicators of $R2 - score$, MAE , MSE and $RMSE$ were sacrificed. From the analysis of specific data, $R2 - score$ decreased by 2.231%; MSE increased by 0.103%; $RMSE$ increased by 1.711%; MAE increased by 1.273%. Comparing BiGRU, CNN-LSTM, LSTNet and DARNN, W-KGLSTM still has advantages.

For GEFCom2012 dataset, the indicators of $R2 - Score$, MSE , $RMSE$ and MAE of KGLSTM are higher than those of BiGRU, CNN-LSTM, LSTNet and DARNN. LSTNet ranks second among the four indicators. Meanwhile, the weighted loss KGLSTM (W-KGLSTM) still shows obvious advantages in peak forecasting, but still sacrifices some indicators. Compared with KGLSTM, the specific data analysis shows that $R2 - score$ decreases by 13.977%; MSE increased by 0.644%; $RMSE$ increased by 5.063%; MAE increased by 4.767%. However, TS number increased by 32%; $Peak - Score$ increased by 13.609%. Compared with LSTNet, WKGLSTM has little difference in $R2 - score$, MSE , $RMSE$ and MAE , but the peak forecasting accuracy is greatly improved. Compared with BiGRU, CNN-LSTM, DARNN. W-KGLSTM has advantages in various indicators.

According to the experimental results of two data sets and four types of comparison models, KGLSTM greatly improves the forecasting accuracy in time series, but there are deficiencies in peak forecasting. Weighted KGLSTM (W-KGLSTM) sacrifices some index accuracy. However, it has obvious advantages in peak forecasting, and its forecasting accuracy is also maintained at a high level.

7 | ABLATION EXPERIMENT

To prove the validity of the model design, careful ablation experiments were carried out. Specifically, one component at a time is removed in the KGLSTM framework. Finally, the weighted KGLSTM and KGLSTM are compared to confirm the importance of the weighted loss function. The KGLSTM with different components removed is named as follows:

- KGLSTMw/oLSTM:

The KGLSTM network model does not have LSTM codec module.

- KGLSTMw/oGCN:

The KGLSTM network model does not have a GCN network module.

- W-KGLSTM:

Use a weighted loss function on the KGLSTM network model.

The comparison of forecasting results of KGLSTMw/oLSTM, KGLSTMw/oGCN, W-KGLSTM and KGLSTM based on two types of data can be shown in Figure. 10 and Figure. 11. The detailed evaluation index comparison table of KGLSTMw/oLSTM and KGLSTMw/oGCN are shown in Table 3. By observing various forecasting charts and index reference tables, it can be seen that:

- The red box in Figure. 10 shows that for the forecasting data of the first 150 days, the forecasting results of KGLSTM is greater than that of W-KGLSTM, and for the forecasting data of the next 150 days. The forecasting results of W-KGLSTM is greater than that of KGLSTM. Meanwhile, for the forecasting data of the first 150 days, forecasting results of KGLSTMw/oGCN on the peak is better than KGLSTMw/oLSTM, but for the forecasting data of the next 150 days, the overall forecasting effect of KGLSTMw/oGCN on the peak is worse than KGLSTMw/oLSTM. This is consistent with the evidence theory in Figure. 8.
- The purple dotted line in Figure. 10 indicates the lowest value of KGLSTMw/oGCN for trough forecasting. Considering that KGLSTMw/oGCN adopts LSTM network structure and bidirectional LSTM encoder, and injects attention mechanism at the same time, in terms of weight, pay attention to the part with higher value, resulting in lower forecasting accuracy for trough.
- The red dotted line in Figure. 10 indicates the lowest value of KGLSTMw/oLSTM for the trough value, because the GCN network will not have the problem of gradient disappearance compared with LSTM during training, and there is no attention injection mechanism. Therefore, compared with KGLSTMw/oGCN, KGLSTMw/oLSTM has higher forecasting accuracy in the trough.
- The green and yellow dotted lines in Figure. 10 represent the lowest value of KGLSTM and W-KGLSTM for trough value. The forecasting of KGLSTMw/oLSTM for wave trough is optimized. in addition, combined with the dynamic weighting method, W-KGLSTM has the best forecasting accuracy in the overall forecasting effect, whether for the forecasting of peak or trough.
- Figure. 11 shows that W-KGLSTM has more outstanding accuracy in load peak forecasting. Compared with KGLSTM, KGLSTMw/oLSTM, KGLSTMw/oGCN, W-KGLSTM has the best peak forecasting effect. However, there is obvious jitter in some time periods, such as the time period around the 250th day. At the same time, RT load forecasting accuracy of W-KGLSTM is also lower than that of KGLSTM. This proves once again that W-KGLSTM sacrifices the overall forecasting accuracy and improves the accuracy of load peak, which shows the effectiveness of dynamic weighting algorithm in peak forecasting.

Table 3 Results of Ablation Experiment

Model	Five year dataset of a city						GEFCom2012 dataset					
	(R2 - Score)	(TS)	(Peak - Score)	(MSE)	(RMSE)	(MAE)	(R2 - Score)	(TS)	(Peak - Score)	(MSE)	(RMSE)	(MAE)
KGLSTMw/oLSTM	0.91939	0.0	0.36775	0.00144	0.03825	0.02948	0.08864	0.0	0.03545	0.04193	0.20478	0.17489
KGLSTMw/oGCN	0.82968	0.0	0.33187	0.00180	0.04245	0.03415	0.50166	0.2	0.32066	0.02293	0.15143	0.13356
W-KGLSTM	0.96763	0.61111	0.75371	0.00148	0.03847	0.02996	0.82850	0.52	0.64340	0.00789	0.08883	0.07824
KGLSTM	0.98994	0.47058	0.67833	0.00045	0.02136	0.01723	0.96827	0.2	0.50731	0.00145	0.03820	0.03057

8 | CONCLUSION

This paper constructs a framework of Edge-Cloud computing energy system, which is used to monitor and analyze the load data collected by edge devices. At the same time, for the task of RT forecasting, a new deep learning framework is proposed. This framework analyzes the characteristics of RT data from the space-time dimension. One branch establishes the relationship between each other from the time dimension, with the time axis as the label and the load data as the characteristic information. One branch establishes mutual dependence from the spatial dimension and takes the house location as the correlation. Taking the five-year load data of a city and the data set of the 2012 global energy forecasting competition as an example, the forecasting of KGLSTM is verified. Experiments indicate that KGLSTM is able to forecast load data of the previous year accurately. Finally, through the weighted processing of KGLSTM, although the weighted KGLSTM sacrifices some experimental accuracy, the weighted KGLSTM can make a higher precision forecasting effect in peak forecasting.

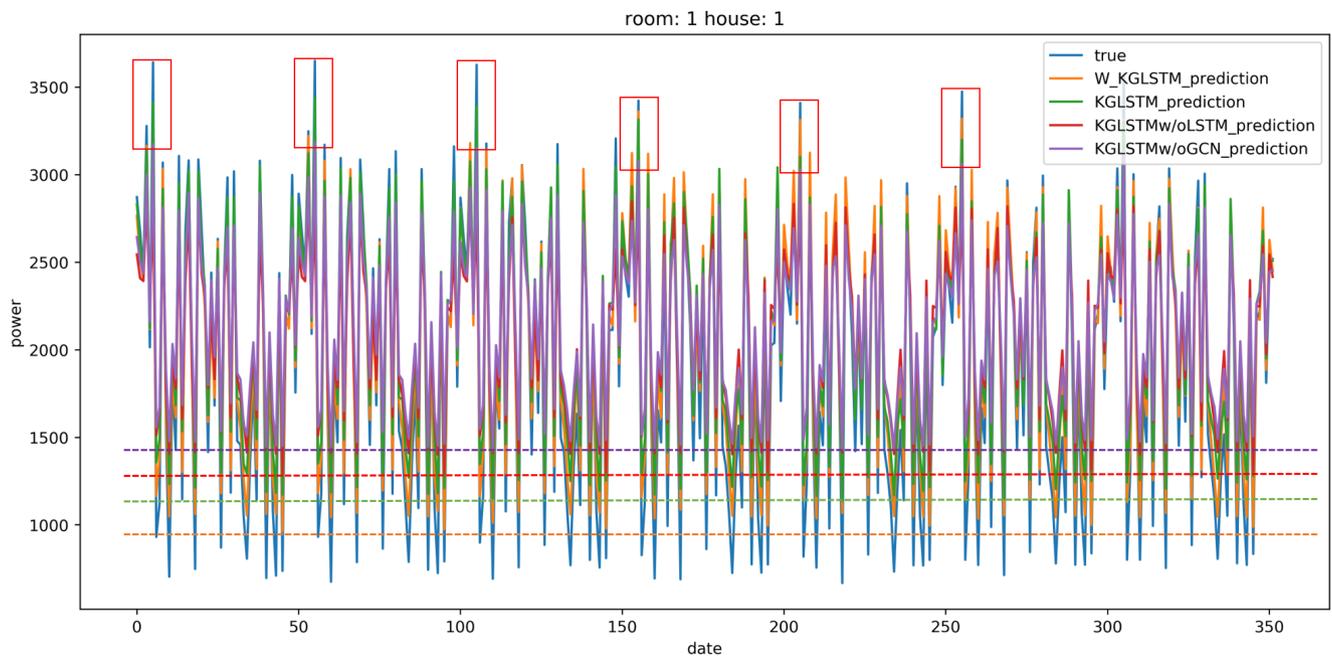


Figure 10 Comparison chart of Ablation Experiment Based on 5-year data set of a city.

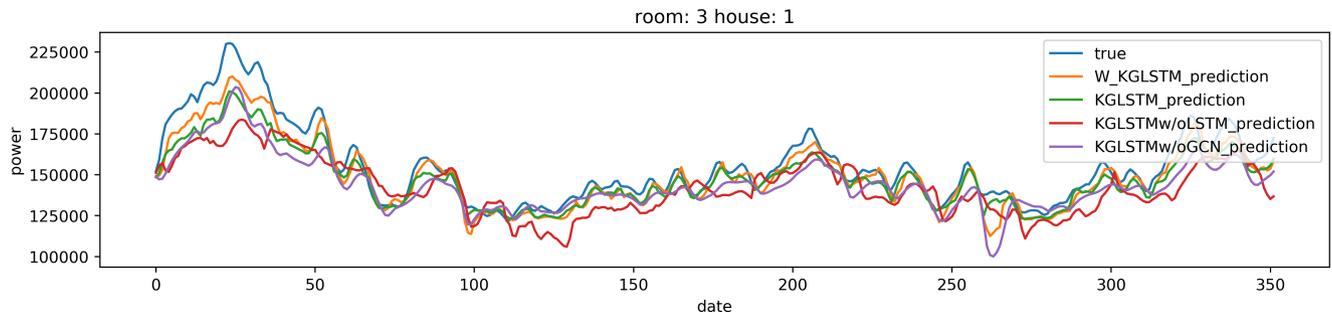


Figure 11 GEFCOM2012 ablation data set comparison.

AUTHOR CONTRIBUTIONS

Qi Liu and Xianming Huang designed the project, oversaw the process, organized and finalized the manuscripts; Jixiang Gan and Qi Liu proposed the methodology, designed the model, drafted manuscript, prepared the data, trained and analyzed the model; Li Pan made the implementation of the computer code and supporting algorithms, and tested the existing code. Xuefei Cao analyzed and evaluated the results. All authors have read and agreed to the published version of the manuscript.

AVAILABILITY OF DATA AND MATERIALS

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

ACKNOWLEDGMENT

This work has received funding from National Key Research and Development Program of China under Grant No.2021YFE0116900, National Natural Science Foundation of China (No. 42275157 and 41975142), Major Program of the National Social Science Fund of China (Grant No. 17ZDA092), and basic Research Programs (Natural Science Foundation) of Jiangsu Province (BK20191398).

References

1. Xu R, Liu Z, Yu Z. Exploring the profitability and efficiency of variable renewable energy in spot electricity market: Uncovering the locational price disadvantages. *Energies* 2019; 12(14): 2820.
2. Liu Q, Kamoto KM, Liu X, Sun M, Linge N. Low-complexity non-intrusive load monitoring using unsupervised learning and generalized appliance models. *IEEE Transactions on Consumer Electronics* 2019; 65(1): 28–37.
3. Liu Q, Nakoty FM, Wu X. A secure edge monitoring approach to unsupervised energy disaggregation using mean shift algorithm in residential buildings. *Computer Communications* 2020; 162: 187–195.
4. Li B, Jia B, Cao W, et al. Application prospect of edge computing in power demand response business. *Power System Technology* 2018; 42(1): 79–87.
5. Xu X, Jiang Q, Zhang P, et al. Game Theory for Distributed IoV Task Offloading With Fuzzy Neural Network in Edge Computing. *IEEE Transactions on Fuzzy Systems* 2022; 30(11): 4593-4604. doi: 10.1109/TFUZZ.2022.3158000
6. Liu Q, Mo R, Xu X, Ma X. Multi-objective resource allocation in mobile edge computing using PAES for Internet of Things. *Wireless Networks* 2020: 1–13.
7. Zhang J, Liu Q, Chen L, Tian Y, Wang J. Nonintrusive Load Management Based on Distributed Edge and Secure Key Agreement. *Wireless Communications and Mobile Computing* 2021; 2021.
8. Yang H, Zhang J, Qiu J, Zhang S, Lai M. A practical pricing approach to smart grid demand response based on load classification. *IEEE Transactions on Smart Grid* 2016; 9(1): 179–190.
9. Tian H, Xu X, Lin T, et al. DIMA: Distributed cooperative microservice caching for internet of things in edge computing by deep reinforcement learning. *World Wide Web* 2022; 25(5): 1769–1792.
10. Xu X, Fang Z, Qi L, Zhang X, He Q, Zhou X. TripRes: Traffic Flow Prediction Driven Resource Reservation for Multimedia IoV with Edge Computing. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 2021; 17(2). doi: 10.1145/3401979
11. Xu X, Fang Z, Zhang J, et al. Edge Content Caching with Deep Spatiotemporal Residual Network for IoV in Smart City. *ACM Trans. Sen. Netw.* 2021; 17(3). doi: 10.1145/3447032
12. Tian H, Xu X, Qi L, et al. CoPace: Edge Computation Offloading and Caching for Self-Driving With Deep Reinforcement Learning. *IEEE Transactions on Vehicular Technology* 2021; 70(12): 13281-13293. doi: 10.1109/TVT.2021.3121096
13. Liu Q, Zhang J, Liu X, et al. Improving wireless indoor non-intrusive load disaggregation using attention-based deep learning networks. *Physical Communication* 2022; 51: 101584.
14. Liu Q, Chen F, Chen F, Wu Z, Liu X, Linge N. Home appliances classification based on multi-feature using ELM. *International Journal of Sensor Networks* 2018; 28(1): 34–42.
15. Li Z, Xu X, Cao X, et al. Integrated CNN and Federated Learning for COVID-19 Detection on Chest X-Ray Images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2022: 1-11. doi: 10.1109/TCBB.2022.3184319
16. Kassahun Y, Gea dJ, Edgington M, Metzen JH, Kirchner F. Accelerating Neuroevolutionary Methods Using a Kalman Filter. In: GECCO '08. Association for Computing Machinery. ; 2008; New York, NY, USA: 1397–1404

17. Yang H, Li W, Luo Cm. Fuzzy adaptive Kalman filter for indoor mobile target positioning with INS/WSN integrated method. *Journal of central south University* 2015; 22(4): 1324-1333.
18. Liu Q, Kamoto KM, Liu X. Microgrids-as-a-service for rural electrification in Sub-Saharan Africa. *Comput. Mater. Contin.* 2020; 63(3): 1249-1261.
19. Lei K, Qin M, Bai B, Zhang G, Yang M. Gcn-gan: A non-linear temporal link prediction model for weighted dynamic networks. In: IEEE. ; 2019: 388-396.
20. Liu Z, Zhu Z, Gao J, Xu C. Forecast methods for time series data: a survey. *IEEE Access* 2021; 9: 91896-91912.
21. Wu H, Liu Q, Liu X, Zhang Y, Yang Z. An edge-assisted cloud framework using a residual concatenate FCN approach to beam correction in the internet of weather radars. *World Wide Web* 2022: 1-27.
22. Salem R, Abdo A. Fixing rules for data cleaning based on conditional functional dependency. *Future Computing and Informatics Journal* 2016; 1(1-2): 10-26.
23. Ilyas IF, Chu X. *Data cleaning*. Morgan & Claypool . 2019.
24. PeiPei Y, Qing L. Kalman filtering of MEMS gyro based on time-series model. In: IEEE. ; 2009: 2-367.
25. Safarinejadian B, Tajeddini MA, Ramezani A. Predict time series using extended, unscented, and cubature Kalman filters based on feed-forward neural network algorithm. In: IEEE. ; 2013: 159-164.
26. Ma Y, Zhang Q, Ding J, Wang Q, Ma J. Short term load forecasting based on iForest-LSTM. In: IEEE. ; 2019: 2278-2282.
27. Duan L, Niu D, Gu Z. Long and medium term power load forecasting with multi-level recursive regression analysis. In: . 1. IEEE. ; 2008: 514-518.
28. Kareem S, Akpinar M. Removing Seasonal Effect on City Based Daily Electricity Load Forecasting with Linear Regression. In: IEEE. ; 2021: 1-6.
29. Li K, Xing Y, Zhu H, Nai W. A short-term hybrid forecasting approach for regional electricity consumption based on Grey Theory and Random Forest. In: IEEE. ; 2020: 194-198.
30. Box GE, Jenkins GM, Reinsel GC, Ljung GM. *Time series analysis: forecasting and control*. John Wiley & Sons . 2015.
31. Ahmed HAY, Mohamed SW. Rainfall Prediction using Multiple Linear Regressions Model. In: IEEE. ; 2020: 1-5.
32. Chen J, Wang J. Prediction of PM2. 5 concentration based on multiple linear regression. In: IEEE. ; 2019: 457-460.
33. Cui C, He M, Di F, Lu Y, Dai Y, Lv F. Research on Power Load Forecasting Method Based on LSTM Model. In: IEEE. ; 2020: 1657-1660.
34. An L, Teng Y, Li C, Dan Q. Combined Grey Model Based on Entropy Weight Method for Long-term Load Forecasting. In: IEEE. ; 2020: 149-153.
35. Vapnik VN. A note on one class of perceptrons. *Automat. Rem. Control* 1964; 25: 821-837.
36. Yan X, Chowdhury NA. Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input. *International Journal of Electrical Power & Energy Systems* 2014; 63: 64-70.
37. Men Dy, Liu Wy. Application of least squares support vector machine (LS-SVM) based on time series in power system monthly load forecasting. In: IEEE. ; 2011: 1-4.
38. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: KDD '16. Association for Computing Machinery. ; 2016; New York, NY, USA: 785-794
39. Wu H, Liu Q, Liu X. A review on deep learning approaches to image classification and object segmentation. *Comput. Mater. Contin* 2019; 60: 575-597.
40. Deng X, Ye A, Zhong J, et al. Bagging-XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Reports* 2022; 8: 8661-8674.

41. Li Z, Lu T, He X, Montillet JP, Tao R. An Improved Cyclic Multi Model-eXtreme Gradient Boosting (CMM-XGBoost) Forecasting Algorithm on the GNSS Vertical Time Series. *Advances in Space Research* 2022.
42. Liu Q, Wu X, Liu X, Zhang Y, Hu Y. Near-Data Prediction Based Speculative Optimization in a Distribution Environment. *Mobile Networks and Applications* 2022: 1-9.
43. Chang YY, Sun FY, Wu YH, Lin SD. A memory-network based solution for multivariate time-series forecasting. *arXiv preprint arXiv:1809.02105* 2018.
44. Roska T, Chua LO. The CNN universal machine: an analogic array computer. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* 1993; 40(3): 163-173.
45. Imani M. Electrical load-temperature CNN for residential load forecasting. *Energy* 2021; 227: 120480.
46. Quan Z, Zeng W, Li X, Liu Y, Yu Y, Yang W. Recurrent neural networks with external addressable long-term and working memory for learning long-term dependences. *IEEE transactions on neural networks and learning systems* 2019; 31(3): 813-826.
47. Yamak PT, Yujian L, Gadosey PK. A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting. In: ACAI 2019. Association for Computing Machinery. ; 2020; New York, NY, USA: 49-55
48. Wang X, Wu J, Liu C, Yang H, Du Y, Niu W. Fault time series prediction based on LSTM recurrent neural network. *journal of beijing university of aeronautics and astronautics* 2018; 44(04): 772-784.
49. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks* 2005; 18(5-6): 602-610.
50. Yang Z, Liu Q, Wu H, Liu X, Zhang Y. CEMA-LSTM: Enhancing Contextual Feature Correlation for Radar Extrapolation Using Fine-Grained Echo Datasets. *Computer Modeling in Engineering & Sciences* 2022.
51. Hossain MS, Mahmood H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *Ieee Access* 2020; 8: 172524-172533.
52. Yang Z, Wu H, Liu Q, Liu X, Zhang Y, Cao X. A self-attention integrated spatiotemporal LSTM approach to edge-radar echo extrapolation in the Internet of Radars. *ISA transactions* 2022.
53. Lai G, Chang WC, Yang Y, Liu H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In: SIGIR '18. Association for Computing Machinery. ; 2018; New York, NY, USA: 95-104.
54. Envelope JYP, Envelope SX, Envelope AJ. Attention-based spatio-temporal graph convolutional network considering external factors for multi-step traffic flow prediction. *Digital Communications and Networks* 2022; 8(3): 343-350.
55. Gao Y, He H, Tan R, Choi J. Combined spatial-temporal energy harvesting and relay selection for cognitive wireless powered networks. *Digital Communications and Networks* 2021; 7(2): 201-213. doi: <https://doi.org/10.1016/j.dcan.2020.07.010>
56. Shih SY, Sun FK, Lee Hy. Temporal pattern attention for multivariate time series forecasting. *Machine Learning* 2019; 108(8): 1421-1441.
57. Wu Z, Pan S, Long G, Jiang J, Chang X, Zhang C. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. In: KDD '20. Association for Computing Machinery. ; 2020; New York, NY, USA: 753-763
58. Qin Y, Song D, Chen H, Cheng W, Jiang G, Cottrell G. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv preprint arXiv:1704.02971* 2017.
59. Razghandi M, Zhou H, Erol-Kantarci M, Turgut D. Short-Term Load Forecasting for Smart Home Appliances with Sequence to Sequence Learning. In: IEEE. ; 2021: 1-6.
60. Xu X, Tian H, Zhang X, Qi L, He Q, Dou W. DisCOV: Distributed COVID-19 Detection on X-Ray Images with Edge-Cloud Collaboration. *IEEE Transactions on Services Computing* 2022: 1-1. doi: [10.1109/TSC.2022.3142265](https://doi.org/10.1109/TSC.2022.3142265)
61. Xu X, Zhu D, Yang X, Wang S, Qi L, Dou W. Concurrent Practical Byzantine Fault Tolerance for Integration of Blockchain and Supply Chain. *ACM Trans. Internet Technol.* 2021; 21(1). doi: [10.1145/3395331](https://doi.org/10.1145/3395331)

62. Liu Q, Kamoto KM, Liu X, et al. A sensory similarities approach to load disaggregation of charging stations in Internet of electric vehicles. *IEEE Sensors Journal* 2020; 21(14): 15895–15903.
63. Liu X, Liu Q, Zou Y, Liu Q. A LSTM-Based Approach to Haze Prediction Using a Self-organizing Single Hidden Layer Scheme. In: Springer. ; 2018: 701–706.
64. She D, Jia M. A BiGRU method for remaining useful life prediction of machinery. *Measurement* 2021; 167: 108277.
65. Zain ZM, Alturki NM. COVID-19 pandemic forecasting using CNN-LSTM: a hybrid approach. *Journal of Control Science and Engineering* 2021; 2021.

