

Cross-Domain Energy Consumption Prediction via ED-LSTM Networks

Ye TAO^{†a)}, Nonmember, Fang KONG[†], Member, Wenjun JU^{††}, Hui LI[†], and Ruichun HOU^{†††}, Nonmembers

SUMMARY As an important type of science and technology service resource, energy consumption data play a vital role in the process of value chain integration between home appliance manufacturers and the state grid. Accurate electricity consumption prediction is essential for demand response programs in smart grid planning. The vast majority of existing prediction algorithms only exploit data belonging to a single domain, *i.e.*, historical electricity load data. However, dependencies and correlations may exist among different domains, such as the regional weather condition and local residential/industrial energy consumption profiles. To take advantage of cross-domain resources, a hybrid energy consumption prediction framework is presented in this paper. This framework combines the long short-term memory model with an encoder-decoder unit (ED-LSTM) to perform sequence-to-sequence forecasting. Extensive experiments are conducted with several of the most commonly used algorithms over integrated cross-domain datasets. The results indicate that the proposed multistep forecasting framework outperforms most of the existing approaches.

key words: cross-domain feature fusion, long short-term memory, encoder-decoder, multistep electricity load forecast

1. Introduction

Electricity is a necessary commodity in contemporary society, and it cannot be stored for future supply. Existing studies show that every increase of 5% in electricity peak demand requires an extra 20% energy output in the absence of an effective electricity load forecasting scheme [1]. Moreover, the increased demand for electricity at certain hours of the day may result in several problems, such as short circuits and transformer failure. To address these issues, many researchers have developed excellent methods and achieved desirable results [2]–[6].

Traditional electricity demand forecasting models are mainly based on data-driven methods and can usually be categorized into (i) statistical learning (SL) models, (ii) machine learning (ML) models, and (iii) deep learning (DL) models. Statistical learning models are mathematical models derived from measurement data and are distinguished by model simplicity and high performance. Examples include the exponential smoothing method [7], the autoregressive integrated moving average (ARIMA) method [8], the seasonal autoregressive integrated moving average (SARIMA)

method [9], etc. However, these models have strict requirements regarding the stability of time series data and involve only univariate data. Furthermore, the rapid deployment of smart meters and building automation systems (BASs) has created opportunities for load forecasting to collect data.

By contrast, ML algorithms do not require strong assumptions about the mapping function and readily learn linear and nonlinear relationships; in other words, they elegantly approximate arbitrary nonlinear functions. Y. Chen et al. [5] proposed a new SVR-based forecasting method, in which determining the SVR model offers a high degree of prediction accuracy and stability in short-term load forecasting. H. Jiang [6] predicted electricity consumption by using support vector regression (SVR) machines. H. Chen et al. [10] proposed an ANN-based short-term load forecasting technique that utilizes the electricity price as one of the main input variables. Compared with SL algorithms, ML models demonstrate remarkable improvements in nonlinear handling ability. However, these ML models feature exceedingly complex engineering modeling and require a certain degree of professional domain knowledge.

Deep learning methods have grown in popularity over the past few years. Many researchers have begun to apply these techniques in the field of electricity load prediction. Deep neural networks (DNNs) are particularly advantageous, as they can automatically learn using a general learning procedure with little prior knowledge and thus do not require domain expertise. Mostafa Askari [11] proposed a load forecasting method by a new composite method based on a multilayer perceptron (MLP) neural network algorithm and obtained excellent results in middle load demand forecasting. L. Li et al. [12] presented a deep convolutional neural network (CNN) model that transforms the numerical prediction problem into an image processing task for electric load forecasting and performs well in terms of accuracy. However, MLP and CNN models only take the current input into account and do not consider a critical factor of the electric data, namely, their time dependency. Consequently, recurrent neural networks (RNNs) have been designed to capture the time dependencies of the previously received input from the current input in the architecture [13]. Rahman et al. [14] applied an LSTM model to make predictions over a time horizon of a few months. Sumit Kumar et al. [15] explored the performance of LSTM and a gated-recurrent unit (GRU) in the field of load forecasting. Kim et al. [16] analyzed the effectiveness of a CNN coupled to an LSTM model to forecast the electricity

Manuscript received November 11, 2020.

Manuscript revised March 16, 2021.

Manuscript publicized May 11, 2021.

[†]The authors are with Qingdao University of Science and Technology, Qingdao, 266071, China.

^{††}The author is with Haier Technology Co., Ltd., 266000, China.

^{†††}The author is with Ocean University of China, 266071, China.

a) E-mail: ye.tao@qust.edu.cn

DOI: 10.1587/transinf.2020BDP0006

consumption of a residence. Furthermore, bidirectional LSTM (Bi-LSTM) [17] and bidirectional GRU (Bi-GRU) [18] have also been studied and achieved higher accuracy than traditional deep learning methods.

While RNNs are capable of extracting the time dependencies from historical data, for the multistep prediction problem, the RNN model and its variant's prediction performance have difficulty reaching the expected target since the sequential temporal dependency of the output label is not considered. Fortunately, a sequence-to-sequence (S2S) RNN model has shown promise in the field of electricity load multistep prediction. It has shown great success in the field of language translation by combining an encoder and a decoder RNN [19].

According to most research, the time series prediction model suffers from shifts introduced by complex factors. However, the majority of existing models aim at the electricity load forecasting of single load data or combine the features associated with load data, such as the voltage and the current. These models might disregard factors associated with seasonal features [20], temporal features [21] and meteorological features [22] in the time series prediction.

This paper presents a short-term multistep prediction framework based on the ED and LSTM models for electricity consumption prediction, named ED-LSTM, that collectively considers cross-domain feature fusion strategies and Deseasonalization (DS) strategies. Cross-domain features such as meteorological conditions, electricity data information and holiday information are combined to calculate the electric load demand through various prediction modules, which aim at preventing the prediction model from becoming affected by complex factor shifts. The DS method is designed to separate the seasonal/trend/residual components of the time series by the seasonal-trend decomposition using LOESS (STL) approach. Then, the residual value is input to the model to predict the load demand. The proposed framework includes three major phases: 1) In the preprocessing phase, typical preprocessing strategies (transformation, DS, normalization, etc.) are applied to the raw data. 2) In the feature fusion phase, features from different domains are extracted and integrated from the input datasets. 3) In the modeling phase, a hybrid LSTM network combined with an encoder-decoder unit is selected to solve the multistep electricity load forecasting problem. Alternative models (MLP, SimpleRNN, GRU, etc.) are also applied to the datasets to evaluate the performance of the proposed method.

The contributions of the paper can be summarized as follows:

A: The self-feature of electricity data exists complex information easily causes the forecasting losses shifts and requires further removing. Seasonal and Trend decomposition using the Loess method (STL) is introduced to tackle the problem. This method extends the study for the data internal feature.

B: Due to the different combinations of domain information derive different performance in same networks, it confirmed that dependencies and correlations exist among

different domains provide important reference information for the accurate prediction of power demand.

C: Compared with four baseline models, the proposed ED-LSTM has an advantage in structure for time series prediction problems. The ED-LSTM model not only inherits the ability of the LSTM model capable of extracting the time dependencies from historical data, but also extracting the output label temporal dependency used encoder-decoder structure. The findings extend and supply the study in short-term multi-step electricity consumption forecasting.

The rest of this paper is organized as follows: Sect. 2 reviews the literature relating to electricity cross-domain features analysis and the LSTM model. Section 3 describes the proposed short-term multistep electricity consumption forecasting framework. Section 4 introduces the electricity dataset and test the accuracy of the proposed ED-LSTM model. Finally, the last section concludes the exposition.

2. Related Work

2.1 Cross-Domain Feature Analysis

Dependencies and correlations exist among different domains provide important reference information for the accurate prediction of power demand, especially those local features closely related to power activities. The weather features are those which have a significant impact on the electricity load demand. Meteorological information is universally used to build the model with the models capable of deducing future electricity load. In many researches, the temperature is most used and considered as effective contribute to the electric load demand [23], [24]. Mayur Barman et al. [25] further studied the temperature, dew point, and wind speed is considered for electricity consumption, founding that temperature and wind speed in a whole have a significant influential to electricity demand. However, only taking into account of meteorological feature is insufficient, the human activity information, namely temporal features, is another factor that affects electricity demand. This is due to the strong temporal regularities of urban life, for example, the workdays would increase the electricity consumption, weekend is inverse. The time series prediction problem of traffic flow is considered the holiday/weekend information as the model input, indicating that consideration of temporal features can decrease the prediction losses [26]. RunHai Jia et al. [27] used the k-means clustering to explore the patterns of manufacturing industrial electricity consume, validating the differs significantly in workdays and holiday/weekend.

Besides meteorological information and temporal information, the self-feature of electricity data exists in complex information, such as trends and seasonal components. Different from the above two domain's data, these components easily cause the forecasting losses shifts and require further removing. Difference method [28] and exponential smoothing [29] was used to remove the impact of seasonal factors. However, these two methods exist limitations that are difficult to cope with long sequences. The Seasonal and

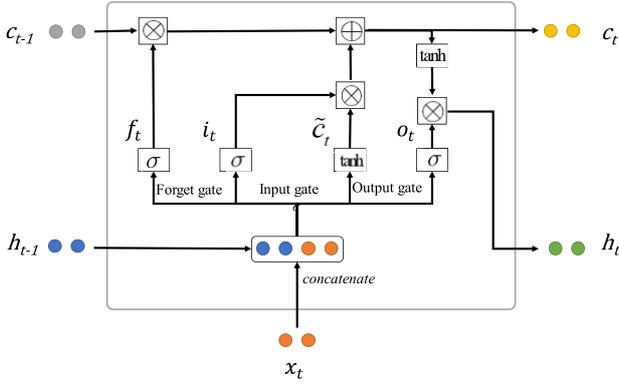


Fig. 1 The LSTM model structure

Trend decomposition using Loess (STL) was proposed to solve the problem in this paper.

In our research, we integrally consider electricity history electricity load information, weather information and temporal information. Further, the data seasonal decompose method is introduced to tackle the error shifts caused by seasonal factor of self-feature in long sequence.

2.2 Long Short-Term Memory Neural Network

Long short-term memory networks (LSTMs) are a variant of RNNs that can effectively learn long-term dependencies from data. Compared with RNNs, LSTMs add an additional cell state unit to transfer the long-term dependency information. The LSTM unit is shown in Fig. 1.

The cell state c_t is a vital part of the LSTM structure; it involves the updating, maintenance and destruction of information in the network. The LSTM units are composed of three gate units (forget gate f_t , input gate i_t , and output gate o_t) to protect and control the cell state. The gates are ways to allow information to pass optionally. The principles of the three gate units can be described with Eqs. (1)–(3):

$$\begin{bmatrix} \tilde{c}_t \\ o_t \\ i_t \\ f_t \end{bmatrix} = \begin{bmatrix} \tanh \\ \sigma \\ \sigma \\ \sigma \end{bmatrix} \left(W \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix} \right) + b, \quad (1)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad (2)$$

$$h_t = o_t \odot \tanh(c_t), \quad (3)$$

where c_t represent the current cell state, \tilde{c}_t represent the new updated cell status values. The notations f_t , i_t and o_t is the sigmoid activation function. c_t represents the internal state of the network, which is specialized for linear circular information transmission and outputs information to the external state h_t of the hidden layer. $x_t \in \mathbb{R}^e$ represents the input at the current moment, $w \in \mathbb{R}^{4d \times (d+e)}$ and $b \in \mathbb{R}^{4d}$ are network parameters.

3. Short-Term Multistep Forecasting Framework

The proposed electricity load forecasting framework consists of 3 modules, data preprocessing, data integration, and

	Power load data	Temporal data	Weather data
t_1	e_1	d_1	w_1
t_2	e_2	d_2	w_2
...
t_n	e_n	d_n	w_n

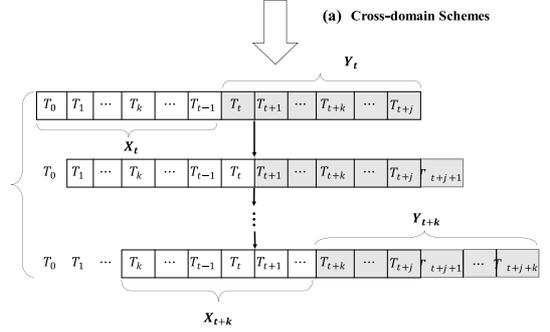


Fig. 2 Data integration module system

model building, as shown in Fig. 3. The de-seasonalization algorithm, cross-domain fusion strategy and temporal feature extraction are designed to obtain highly accurate prediction results.

3.1 Data Pre-Processing Module

The main function of the data preparation module is to perform the extract, transform and load (ETL) process for three sources of data. The purpose of the ETL process is to integrate scattered, disordered, and inconsistent data from different domains to provide an analytical basis for subsequent electricity forecasting.

Since machine learning models only take digital data as input, a label encoder is used to encode the classified data. As the dimensionalities of the features may be different, to avoid the negative effects of weight updating during the training process, normalization is required to obtain dimensionless expressions. Therefore, the indexes of different units can be compared and weighed.

As a branch of time series forecasting problem, the history load data usually contains 3 components: trend T , seasonal S , and residual R . Seasonality and Trend components lead to high volatility of serial data, which disturbs forecasting accuracy [30]. Therefore, the seasonal decompose method is used to extract residual component R as the input data of model, as shown in Eqs. (4)–(6):

$$Y_t = T_t + S_t + R_t, \quad (4)$$

$$T_t = \left(X_{t-f-1/2} + X_{t-f-1/2+1} + \dots + X_{t+f-1/2} \right) f^{-1}, \quad (5)$$

$$S_t = \left(0.5 * X_{t-f/2} + X_{t-f/2+1} + \dots + X_{t+f/2-1} + 0.5 * X_{t+f/2} \right) f^{-1}, \quad (6)$$

$$\hat{S}_t = Y_t - T_t, \quad (7)$$

$$S_t = \sum_{i=0}^n f^{-1} \hat{S}_{t+f*i}, \quad (8)$$

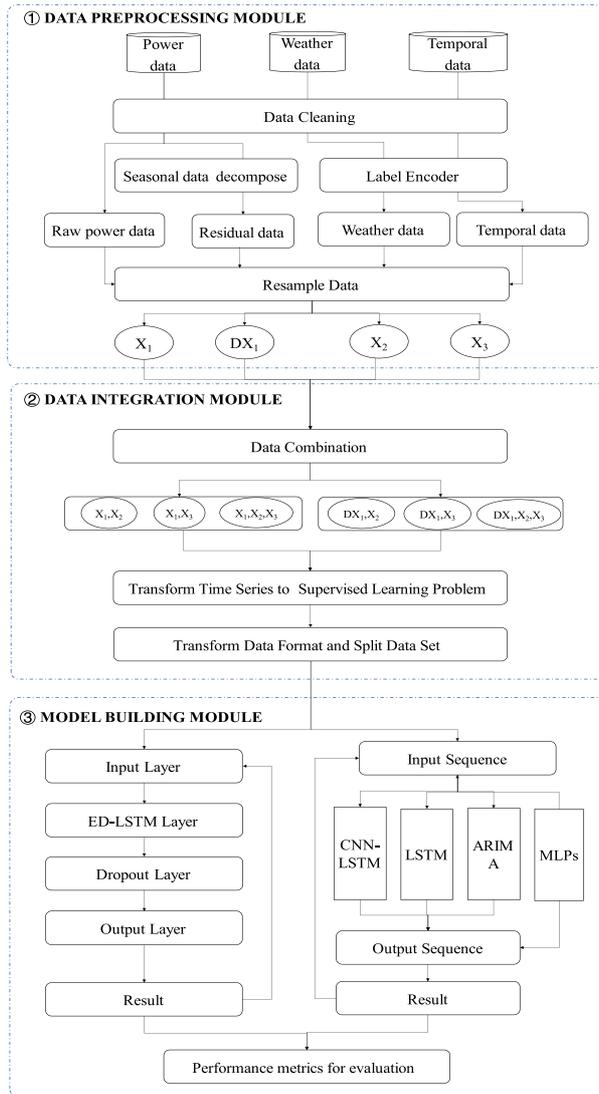


Fig. 3 The proposed electricity load forecasting system

$$R_t = Y_t - S_t - T_t, \quad (9)$$

where f is the frequency and l are the length of time series. When t is even, using the Eq. (5) calculate the trend T_t , otherwise, use Eq. (6) calculate the T_t .

3.2 Data Feature Integration Module

Although the self-features of history electricity load data provide meaningful information for prediction models, the impact of temporal information, namely daily periodicity and holiday/weekend, are to be not overlooked. In addition, weather information would cause error shifts that affect the patterns of load forecasting. Therefore, it may increase losses if only the history electricity information is taken into consideration without referring other domain information. To solve this problem, the proposed forecasting framework introduces the feature integration module that, including the cross-domain data fusion method and refining the time series problem to a supervised learning problem.

3.2.1 Cross-domain Data Fusion Method

Traditional electricity consumption forecasting usually focuses on a single domain. However, with the continuous development of big data technologies, diversity and heterogeneity of datasets are shown in electricity forecasting subjects from different sources in a different domain. These datasets are likely to be composed of more fine-grained patterns, such as temporal attribute, temp feature and dew-point feature, *etc.* How to explore the relevance of the different disparate domains is predominant, essentially distinguishing cross-domain forecasting tasks from single data source forecasting tasks. Therefore, advanced methods that can integrate data from multiple domains into a machine learning model are needed. So, we proposed a cross-domain data fusion strategy that merges incorporate multiple temporal features and weather features to predict electricity consumption, as showed in Fig. 2 (a).

In Fig. 2, $T = \{d_0, d_1, \dots, d_n\}$, $E = \{e_0, e_1, \dots, e_n\}$, $W = \{w_0, w_1, \dots, w_n\}$ represent temporal feature vector data, weather feature vector data and electricity load data, respectively. The first is to merge these data T , E , W into S , and the next is to map the multiple domain data to a data frame.

3.2.2 The Time Series Problem into the Supervised Learning Problem

A supervised learning problem consists of the input feature set X and the output label set Y that an algorithm can learn how to predict the output sequence from the input sequence. However, the original electricity dataset is a kind of continuous sequence sorted by time index, it cannot be directly provided to the machine learning model for training, so it has to be redefined as a supervised learning problem. We

Algorithm 1 The Data Integration Algorithms

Input: Raw multiple domain data of: $L \in \mathbb{R}^l$, $M \in \mathbb{R}^m$,

Output: Feature Sample $X \in \mathbb{R}^{l+m+t}$, Label Sample $Y \in \mathbb{R}^{l+m+t}$

- 1: resample L, M, F to align timestamps
 - 2: $META \leftarrow \text{concat}(L, M, F)$
 - 3: $META \leftarrow \text{normalize}(META)$
 - 4: initial $s \leftarrow 0$
 - 5: initial $w \leftarrow 0$
 - 6: define empty list of $X \in \mathbb{R}^{l+m+t}$ and $Y \in \mathbb{R}^{l+m+t}$
 - 7: for $i = 0$ to length ($META$) do
 - 8: $i \leftarrow s + w$
 - 9: $j \leftarrow i + w$
 - 10: if $i < \text{length}(META)$ then
 - 11: $X \leftarrow \text{concatenate}(X, META_{(s:i)})$
 - 12: $Y \leftarrow \text{concatenate}(Y, META_{(i:j)})$
 - 13: $s \leftarrow s + 1$
 - 14: end for
 - 15: return X, Y
-

design Algorithm 1 to transform a time series problem into a supervised learning problem. The main idea of algorithm 1 is utilize the sliding window algorithm to divide the original time series data D into feature set X and label set Y by specifying the step size s and the window size w , as shown in Fig. 2 (b).

3.3 Encoder-Decoder Long Short-Term Memory (ED-LSTM) Model

The framework of the proposed ED-LSTM is shown in Fig. 4. The LSTM unit is employed as the encoder and decoder scheme to improve the learning of the continuity in the input and output sequences. Given an input sequence $x_{1:S}$, another output sequence $y_{1:T}$ is generated by the ED-LSTM model. In the encoder phase, the input sequence $x_{1:S}$ is then updated by Eqs. (1)–(3) in the LSTM unit to generate the state unit containing the input sequence summary. Then, a LSTM network f_{enc} is used to encode the input sequence $x_{1:S}$ to obtain a fixed dimension vector C , which is the hidden state at the last unit of the LSTM. In the decoder phase, as the target sequence $y_{1:T}$ is generated, another LSTM neural network is used for decoding, assuming that the prefix sequence $y_{1:T}$ is generated at the t step. The decoder receives the vector C as the initial cell state for the sequence. The decoding step is initiated with a dummy input s_0 . Then, the decoder recursively decodes the vector to generate the output sequence. It then feeds the output values obtained in the previous update as the current update's input values. The number of reused LSTM units in the encoding and decoding phases depends on the input and output sequence lengths. Finally, the output sequence $\{y_1, y_2, y_3\}$, whose length is arbitrary, is produced recursively using the ED-LSTM model, as shown in Eqs. (10)–(15).

$$h_t^e = f_{enc}(h_{t-1}^e, e_{x_{t-1}}, \theta_{enc}), \forall t \in 1 : S, \quad (10)$$

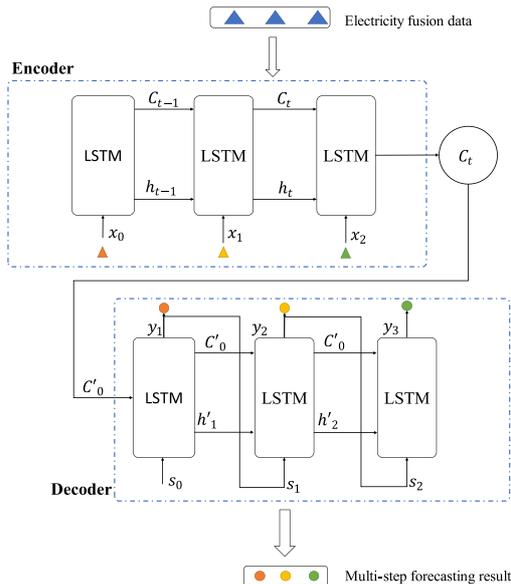


Fig. 4 The ED-LSTM model structure

$$u = h_s^e, \quad (11)$$

$$h_0^d = C_t, \quad (12)$$

$$h_t^d = f_{dec} h_{t-1}^d, y_{t-1}, \theta_{dec}, \quad (13)$$

$$o_t = g h_t^d, \theta_o, \quad (14)$$

$$y_t = f_{dec} y_{t-1}, s_t, h_t^d, \quad (15)$$

4. Results and Discussion

4.1 Dataset Description

In this paper, we use a multiple feature aggregate dataset to validate the performance of the proposed electricity load forecasting architecture, including electricity load data, weekend/holiday data and meteorological data. The first dataset is the total electric load consumption dataset collected from different manufacturers. The second is the meteorological dataset collected from the climatological station located in the manufacturers' home regions. The last is the temporal feature data, such as weekend and holiday data. The raw time resolution of the electricity load dataset is 5 minutes. To analyze the ability of the dataset to forecast the hourly electricity demand, the dataset is resampled. Detailed information on the datasets is described in Table 1. The domain data features correlated with electricity consumption are depicted in Table 2. The training set consists of the first eight years of data, and the ninth year's data are used as the testing set. For the dataset, we select 80% of the data as the training data and 20% of the data as the testing data.

The hardware and software environment of the experiment is provided as follows, including the configuration of relevant parameters. The Keras open-source underlying framework based on TensorFlow 2.4 is used to establish the deep learning model, and Scikit-learn is used to

Table 1 Dataset description

Dataset	Load/Meteorological/Temporal
Time interval	2010/01/01-2014/12/31
Time span	1 hours
Feature num	1/3/4
Total sample	43824

Table 2 Related domain feature description

Feature	Code values	Description
Load	[0-5932]	the range of electricity load
Temperature	[6-102]	the range of temperature
Dew point	[0-52]	the range of dew point
Wind speed	[-22-78]	the range of wind speed
Dow	0-7	Day of the week
Month	1-12	Month of the year
Weekend	0,1	0: non-weekend; 1: weekend
Holiday	0,1	0: non-holiday; 1: holiday

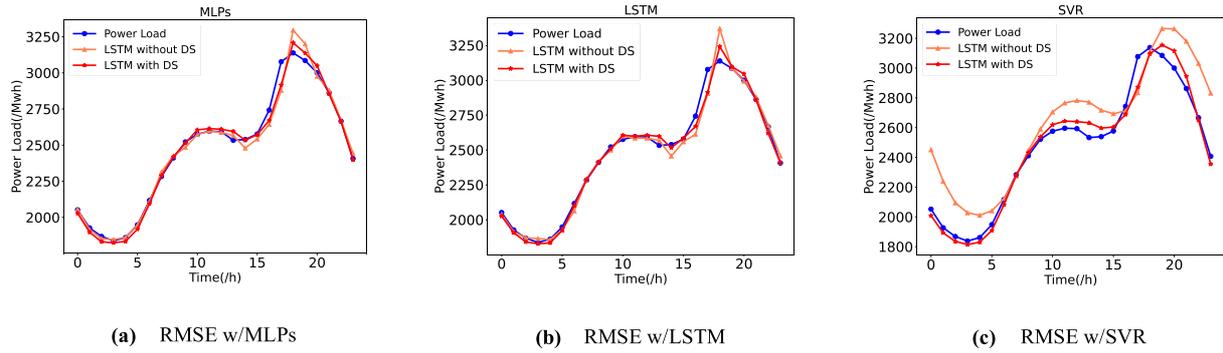


Fig. 5 The prediction results of MLPs/SVR/LSTM between original electricity dataset and de-seasonalization electricity dataset

build ARIMA/SARIMA models. All experiments are conducted on an ubuntu server with an Intel(R) Xeon(R) Silver 4210 CPU @ 2.20 GHz configuration, 4 GPUs, each with a 12 GB GeForce RTX 2080 Ti, and 64 GB memory.

4.2 Evaluation Criteria

The prediction results are compared by three evaluation metrics: the root mean square error (RMSE) [31], coefficient of determination (R^2), and mean absolute percentage error (MAPE). These three indexes reflect that the overall prediction error intuitively varies according the target mean [33]. The RMSE and MAPE values constantly decrease during model training, with smaller values implying a better model fit to the data. Conversely, the closer the value of R^2 is to 1, the better the linear regression fits the data. The detailed equations of the three metrics are formulated in Eqs. (16)–(18)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}, \quad (16)$$

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right|, \quad (17)$$

$$R^2 = 1 - \frac{\sum_i \hat{y}_i - y_i^2}{\sum_i \bar{y}_i - y_i^2}, \quad (18)$$

4.3 Deseasonalization Effect Comparison

Figure 5 shows the prediction performance of the three models after deseasonalization. In general, the DS method decreases the prediction error rate of the model, and the SVR model most substantially reflects the influence of seasonal factors on electricity consumption forecasting. The results indicate the effectiveness of the DS method in electrical load forecasting. In each subgraph, the peak forecasting values of the power load reflect the substantial impact of the seasonal factors. Adjusting for these factors, the LSTM, MLP, and SVR models all improved in predicting the peak and trough

values.

According to the experimental results, the DS approach based on the patterns of the seasonal periods enhances the forecasting performance of nonstationary time series with trend and seasonal components.

4.4 Cross-Domain Feature Impact Comparison

To analyze the impact of different domain information on electricity consumption forecasting, combinations of various features, including load-self features (L), temporal features (T), and weather features (W), were used. These combinations are fed into different kinds of RNN models with different feature combinations. The experimental results are visually presented in Fig. 6.

Overall, as the number of features increases, the RMSE decreases, which indicates that the combination of different domain attributes is conducive to improving energy load forecasting, as shown in the six subgraphs in Fig. 6. For subgraph (a), the load data features (L) are combined with the W and T features, and the error rates of the W- and T-based models are lower than those of the model created with L alone, showing that the model utilizing cross-domain features is superior to the single feature-based model. Moreover, the model created with feature W has a slightly lower RMSE than the model created with feature T, demonstrating that energy load consumption forecasting is more sensitive to weather features than temporal features. Furthermore, compared with the model constructed with the combinations of L + T and L + W, the model constructed with L, T, and W yielded the best forecasting result. The results indicate that energy load consumption is related to holiday factors and the weather condition.

Figure 6 shows the error rate of the LSTM model and other baseline models. It can be observed that the LSTM model outperforms the four baseline models with the cross-domain feature fusion dataset. This indicates not only the advantage of taking multiple features into consideration over individual features but also that the LSTM model is better at extracting complex patterns from the various domain features than the reference models.

In general, each algorithm shows a significant dif-

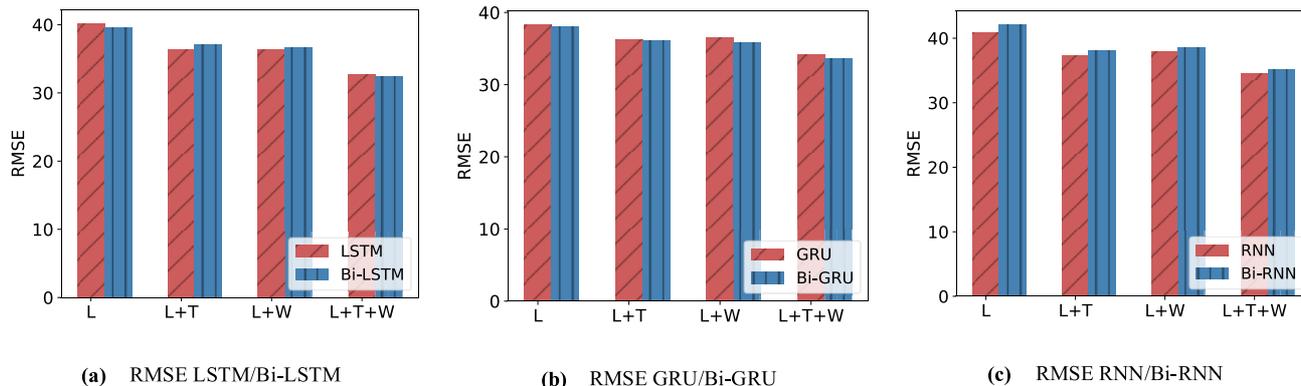


Fig. 6 The RMSE values of RNNs model using cross-domain data fusion method

Table 3 The error values of electricity forecasting models-based L + T + W

	RNN	Bi-RNN	LSTM	Bi-LSTM	GRU	Bi-GRU
RMSE	35.16	34.28	33.88	33.52	35.21	34.74
R ² (%)	99.63	99.65	99.68	99.68	99.61	99.65
MAPE(%)	1.06	1.04	1.01	1.02	1.12	1.07

ference for different feature combinations. RNN/Bi-RNN and LSTM/Bi-LSTM are more sensitive to temporal characteristics than weather characteristics. For the GRU/Bi-GRU algorithm, weather features improve the prediction result more than time series features. Table 3 shows the predictive performance of the LSTM and baseline models. It can be seen that the LSTM/Bi-LSTM model outperforms the four baseline models. Compared with the RNN/Bi-RNN, GRU/Bi-GRU and Bi-LSTM models, the reductions of the LSTM model in RMSE are respectively 3.64%/1.17%, 3.78%/2.47%. and -1.07%. The RMSE of the LSTM model is slightly inferior to that of Bi-LSTM, ranking second. However, compared with the LSTM model, the Bi-LSTM has a complex structure and long calculation time. Therefore, we choose LSTM as the benchmark of the encoder-decoder model.

The experimental results show that the error values decrease as the number of features increase, suggesting that the proposed cross-domain feature fusion method has more accurate forecasting results than the single feature forecasting model in the electricity consumption prediction. Furthermore, compared with the variant of RNNs model, the LSTM model, which takes additional potential temporal information and weather information into account for electricity load demand prediction, is beneficial for improving predictive performance.

4.5 Model Comparison Based Multi-Step Forecasting

To evaluate the effectiveness of the proposed multistep forecasting model, four baseline models are selected and compared with the ED-LSTM model: the ARIMA/SARIMA

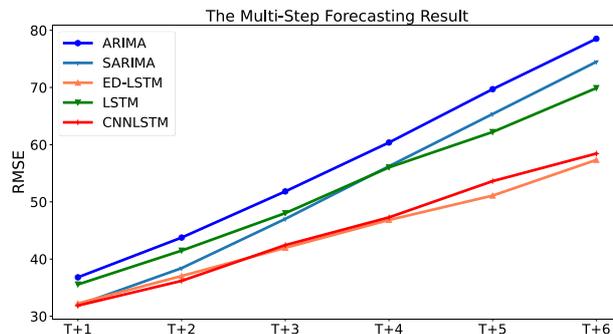


Fig. 7 The multi-step forecast result of ED-LSTM/ARIMA/SARIMA/LSTM/CNN-LSTM on cross-domain electricity consumption dataset (L + W + T)

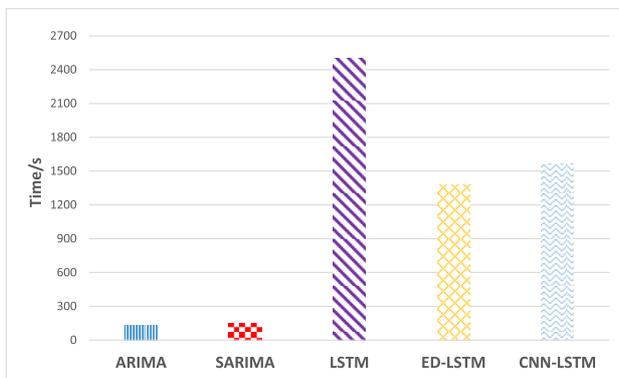
model, the LSTM model, and the latest CNN-LSTM model.

Figure 7 presents the predictive performance of the multistep forecasting framework for the next six steps. Overall, the ARIMA algorithm has the highest prediction error, while the ED-LSTM and CNN-LSTM algorithms show the lowest prediction performance, as well as similar prediction performance. Over time, the prediction error of the ED-LSTM algorithm accumulates slower than the LSTM algorithm, which indicates that the prediction error can be reduced by considering the temporal dependence of the output labels.

Table 4 shows the predictive error values of the proposed ED-LSTM model and the baseline models. It can be observed that the RMSE of the ED-LSTM model is superior to that of the other baseline models over more than half the steps. Compared ED-LSTM with the single-feature

Table 4 The RMSE values of multi-step forecasting of electricity load based on cross-domain feature fusion and DS methods

Forecast step	ARIMA	SARIMA	ED-LSTM	LSTM	CNN-LSTM
T + 1	36.81	31.95	32.24	35.56	31.87
T + 2	43.76	38.41	37.04	41.48	36.19
T + 3	51.83	47.02	41.96	48.03	42.44
T + 4	60.38	56.23	46.83	56.03	47.29
T + 5	69.71	65.37	51.1	62.22	53.63
T + 6	78.52	74.49	57.84	69.9	58.45

**Fig. 8** Comparison of different models in execution efficiency

ARIMA/SARIMA algorithms, the prediction error of ED-LSTM method is lower in all time steps. This validates the advantage of implementing cross-domain feature fusion strategies over individual load features as the model training data. This also shows that cross-domain characteristic data can reduce the influence of uncontrollable factors on power prediction. Compared with the multifeatured LSTM model, the ED-LSTM model is superior for most time steps. Furthermore, it can be seen that accounting for the temporal dependence of the output labels can improve the prediction performance for many-to-many prediction problems. Compared with that of the CNN-LSTM model, the RMSE of the ED-LSTM model is slightly higher in the first two-time steps, but in the last four steps, the RMSE of ED-LSTM is lower. It indicates that the accuracy of ED-LSTM is better than the CNN-LSTM in long forecasting horizons. This conclusion again illustrates the importance of considering cross-domain feature data and the temporal dependence of the output sequence tags.

Figure 8 depicts the computational efficiency of the proposed model and baseline models in terms of running time. It can be seen that the ARIMA/SARIMA model has the shortest running time among all models. This shows that the statistical model is simple in structure and fast in calculation speed. The CNN-LSTM model requires the longer running time than ED-LSTM model, indicating that the combination of the CNN model and LSTM model increases the complexity and depth of the training data calculations.

The ED-LSTM model has no obvious advantages over the SARIMA model in terms of operation efficiency, but the running time is within an acceptable range compared with the CNN-LSTM model and LSTM model. This shows that the ED-LSTM model strikes a balance between prediction performance and operation efficiency.

5. Conclusion

In this paper, a sequence-to-sequence multistep prediction framework is proposed. This framework combines encoder-decoder and LSTM models and uses seasonal adjustment strategies and cross-domain feature aggregation strategies to improve prediction performance. Relying on the seasonal adjustment strategy, the residual of the time series is decomposed to make the prediction, which improves the prediction ability of the model at the peak value. Furthermore, cross-domain feature fusion ($T + W + L$) is used to generate training samples, and the prediction performance of the resulting is greater than that of the single-feature models, indicating that the multifeature fusion method has certain advantages. Moreover, the ED-LSTM model is based on the RNN model, which is used to extract the time-dependent characteristics of the input data, increasing the ability to capture the time-dependent characteristics from the output tags and further improving the prediction performance.

In terms of the prediction horizons, although the prediction accuracy of the proposed multistep prediction scheme decreases as the number of prediction steps increases, the rate of decline rate is obviously lower than that of the ARIMA/SARIMA, and LSTM models and slightly lower than that of the CNN-LSTM model. In terms of the predicted performance, the ED-LSTM model is not the most suitable but is obviously better than the CNN-LSTM model. This conclusion shows that in the field of short-term multistep forecasting, the proposed load forecasting framework not only captures the temporally dependent characteristics between the input and output data but also avoids the influence of complex factors due to active seasonal and artificial features. Compared with the LSTM model, the ARIMA/SARIMA, and CNN-LSTM models have certain advantages. These findings extend and complement the research on the influencing factors and horizons of power load forecasting. In future work, we aim to include more cross-

domain resources, e.g., to include household power consumption, so that the working state of an individual household or single piece of electrical equipment can be accurately predicted.

Acknowledgments

This work was partly supported by National Key Research and Development Plan under Grant 2017YFB1400903 and Shandong Province Colleges and Universities Young Talents Initiation Program under Grant 2019KJN047.

References

- [1] S. Gyamfi and S. Krumdieck, "Scenario analysis of residential demand response at network peak periods," *Electric Power Systems Research*, vol.93, pp.32–38, 2012.
- [2] X. Shao, C.-S. Kim, and P. Sontakke, "Accurate deep model for electricity consumption forecasting using multi-channel and multi-scale feature fusion CNN-LSTM," *Energies*, vol.13, no.8, p.1881, 2020.
- [3] A.K. Fard and M.-R. Akbari-Zadeh, "A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting," *Journal of Experimental & Theoretical Artificial Intelligence*, vol.26, no.2, pp.167–182, 2014.
- [4] W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol.10, no.1, pp.841–851, 2017.
- [5] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol.195, pp.659–670, 2017.
- [6] H. Jiang, Y. Zhang, E. Muljadi, J.J. Zhang, and D.W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Trans. Smart Grid*, vol.9, no.4, pp.3341–3350, 2016.
- [7] E. Ostertagova and O. Ostertag, "Forecasting using simple exponential smoothing method," *Acta Electrotechnica et Informatica*, vol.12, no.3, pp.62–66, 2012.
- [8] H. Nie, G. Liu, X. Liu, and Y. Wang, "Hybrid of ARIMA and SVMs for short-term load forecasting," *Energy Procedia*, vol.16, pp.1455–1460, 2012.
- [9] D. Chikobvu and C. Sigauke, "Regression-SARIMA modelling of daily peak electricity demand in South Africa," *Journal of Energy in Southern Africa*, vol.23, no.3, pp.23–30, 2012.
- [10] H. Chen, C.A. Canizares, and A. Singh, "ANN-based short-term load forecasting in electricity markets," 2001 IEEE power engineering society winter meeting. Conference proceedings (Cat. No. 01CH37194), vol.2, pp.411–415, IEEE, 2001.
- [11] M. Askari and F. Keynia, "Mid-term electricity load forecasting by a new composite method based on optimal learning MLP algorithm," *IET Generation, Transmission & Distribution*, vol.14, no.5, pp.845–852, 2019.
- [12] X. Dong, L. Qian, and L. Huang, "Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach," 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), pp.119–125, IEEE, 2017.
- [13] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol.404, p.132306, 2020.
- [14] A. Rahman, V. Srikumar, and A.D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Applied energy*, vol.212, pp.372–385, 2018.
- [15] S. Kumar, L. Hussain, S. Banarjee, and M. Reza, "Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster," 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), pp.1–4, IEEE, 2018.
- [16] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol.182, pp.72–81, 2019.
- [17] S. Wang, X. Wang, S. Wang, and D. Wang, "Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol.109, pp.470–479, 2019.
- [18] M. Jaihumi, J.K. Basak, F. Khan, F.G. Okyere, E. Arulmozhi, A. Bhujel, J. Park, L.D. Hyun, and H.T. Kim, "A Partially Amended Hybrid Bi-GRU—ARIMA Model (PAHM) for Predicting Solar Irradiance in Short and Very-Short Terms," *Energies*, vol.13, no.2, p.435, 2020.
- [19] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.
- [20] G. Dudek, "Forecasting time series with multiple seasonal cycles using neural networks with local learning," *International Conference on Artificial Intelligence and Soft Computing*, pp.52–63, Springer, 2013.
- [21] D. Yang et al., "MF-CNN: Traffic flow prediction using convolutional neural network and multi-features fusion," vol.102, no.8, pp.1526–1536, 2019.
- [22] M.G. Fikru and L. Gautier, "The impact of weather variation on energy consumption in residential houses," *Applied Energy*, vol.144, pp.19–30, 2015.
- [23] P. Mandal, T. Senjyu, N. Urasaki, and T. Funabashi, "A neural network based several-hour-ahead electric load forecasting using similar days approach," *International Journal of Electrical Power & Energy Systems*, vol.28, no.6, pp.367–373, 2006.
- [24] M. Barman, N.D. Choudhury, and S. Sutradhar, "A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India," *Energy*, vol.145, pp.710–720, 2018.
- [25] M. Barman and N.B.D. Choudhury, "Season specific approach for short-term load forecasting based on hybrid FA-SVM and similarity concept," *Energy*, vol.174, pp.886–896, 2019.
- [26] D. Yang, S. Li, Z. Peng, P. Wang, J. Wang, and H. Yang, "MF-CNN: Traffic flow prediction using convolutional neural network and multi-features fusion," *IEICE Trans. Inf. & Syst.*, vol.E102-D, no.8, pp.1526–1536, 2019.
- [27] R. Jiao, T. Zhang, Y. Jiang, and H. He, "Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network," *IEEE Access*, vol.6, pp.59438–59448, 2018.
- [28] S. Bouktif, A. Fiaz, A. Ouni, and M.A. Serhani, "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting," *Energies*, vol.13, no.2, p.391, 2020.
- [29] S. Bouktif, A. Fiaz, A. Ouni, and M.A. Serhani, "Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol.11, no.7, p.1636, 2018.
- [30] X. Tang, Y. Dai, T. Wang, and Y. Chen, "Short-term power load forecasting based on multi-layer bidirectional recurrent neural network," *IET Generation, Transmission & Distribution*, vol.13, no.17, pp.3847–3854, 2019.
- [31] R.G. Pontius, O. Thonteh, and H. Chen, "Components of information for multiple resolution comparison between maps that share a real variable," *Environmental and Ecological Statistics*, vol.15, no.2, pp.111–142, 2008.
- [32] N.R. Draper and H. Smith, *Applied regression analysis*, John Wiley & Sons, 1998.
- [33] R.J. Hyndman and A.B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol.22, no.4, pp.679–688, 2006.



Ye Tao received the B.S. and Ph.D. degrees from the Department of Computer Science and Technology, Shandong University, in 2005 and 2009, respectively. He is currently a professor with the College of Information Science and Technology, Qingdao University of Science and Technology. His research interests include service computing and software engineering.



Fang Kong received the B.S. degree, in 2018. He is currently the master's degree with the Qingdao University of Science and Technology. His research interests include data fusion and time series analysis.



Wenjun Ju received the Ph.D. degree from Southwest Jiaotong University in 2005. He is now the technical director of Haier U+ platform. His research interests include smart home, value-chain integration and data mining.



Hui Li received the Ph.D. degree from Wuhan University of Technology in 2013. He is currently an associated professor with the College of Information Science and Technology, Qingdao University of Science and Technology. His research interests include deep learning techniques and applications.



Ruichun Hou is now a senior engineer with Ocean University of China. Her research interests include manufacturing informatization and intelligent manufacturing system.