Electrical Power and Energy Systems 135 (2022) 107517

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Multi-task short-term reactive and active load forecasting method based on attention-LSTM model☆

Jiaqi Qin [a], Yi Zhang [a], Shixiong Fan [b], Xiaonan Hu [c], Yongqiang Huang [a], Zexin Lu [a], Yan Liu [d,*]

[a] College of Computer Science, Sichuan University, Chengdu 610000, China
[b] China Electric Power Research Institute, Beijing 100192, China
[c] College of Computer Science & Software Engineering, Shenzhen University, Shenzhen 518060, China
[d] College of Electrical Engineering, Sichuan University, Chengdu 610000, China

ARTICLE INFO

Keywords:
Active load forecasting
Reactive load forecasting
Multi-task learning
Attention mechanism
Long short-term memory
Grid search

ABSTRACT

With the rapid development of power markets, smart grids and large-scale renewable energy generation, it is crucial to be able to accurately predict both reactive and active power loads. In this paper, we propose a novel multi-task load forecasting method for predicting both active and reactive power simultaneously. The long short-term memory (LSTM) architecture is employed in the backbone prediction model, supported by an attention mechanism to prevent performance deterioration. Considering the latent dynamic correlations between the reactive and active loads of a substation, multi-task regression based on hard parameter sharing is adopted to treat the forecasting of both types of loads as parallel subtasks. Meanwhile, we design an adaptive combined-task-wise loss function to optimize the proposed multi-task load forecasting model to avoid biasing of the final model for any subtask. We compare our multi-task attention-LSTM (MTAL) model with other popular single-task load forecasting models and achieve superior accuracy on both subtasks. The results indicate that the proposed method is robust and reliable for practical applications in power systems.

1. Introduction

Power load forecasting, in which the aim is to predict future power loads based on historical power loads, has become a topic of great interest in both research and engineering fields in the past two decades. Accurate load forecasting, especially short-term load forecasting, is critical for improving the utilization rate of electric generator devices and the effectiveness of economic dispatch as well as optimizing energy management [1].

There are two different kinds of power in an alternating current (AC) power transmission system: active power and reactive power. Active power is converted into other forms of energy, such as mechanical energy and luminous energy, and is consumed directly by users. Reactive power, which is used to establish a magnetic field for the purpose of exchanging energy, continuously flows along the transmission lines and cannot be neglected due to the reactance of the lines, transformers and user devices in the AC transmission system [2]. With the development of smart grids, large-scale renewable energy generation and the forthcoming reactive power market, reactive load forecasting is becoming increasingly important for the following reasons. First, many important operations in power transmission systems are based on analyses of reactive power, such as voltage and reactive power optimization, reactive power elimination and quality assurance for electric power [3–5]. Second, due to advances in smart grid technology, the demand side is preferred to be responsible for equalizing generation with demand. Thus, a smart power transmission network requires accurate reactive power forecasting to optimize energy dispatch and ensure the stable operation of the grid [6]. Third, with the rapid development of large-scale renewable energy generation, the architecture of power generation systems has changed from being centralized to being distributed in nature. As a result, the uncertainty of the reactive power in the grid has increased [7]. Last but not least, the theory of reactive power markets indicates that treating reactive power as merchandise will play an important role in encouraging the supply of reactive power while guiding rational reactive power investment. In this context, precise reactive load forecasting will be a key step for price establishment

☆ This work was supported in part by the National Natural Science Foundation of China (61902264) and in part by the Key R&D Program of Sichuan Province (2019YFS0125).

* Corresponding author.
E-mail address: liuyan77@scu.edu.cn (Y. Liu).

https://doi.org/10.1016/j.ijepes.2021.107517
Received 31 May 2020; Received in revised form 9 August 2021; Accepted 14 August 2021
Available online 27 August 2021

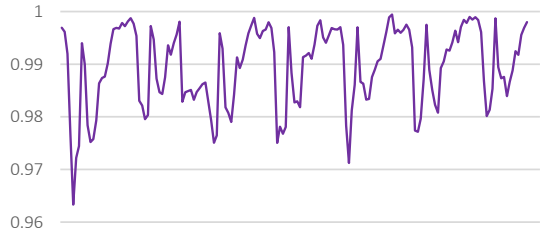


Fig. 1. Power factor curve (December 3nd-9th, 2018).

[8]. In summary, there is an urgent need for accurate forecasting of both active and reactive loads, and this situation has inspired us to develop a proper forecasting solution.

Since the relationship between the active and reactive loads is constrained by the power factor, the reactive load can be roughly estimated from the active load. Normally, the power factor is calculated as the ratio of the active power to the apparent power. However, as shown in Fig. 1, the power factor is not a constant value but varies over time. Thus, it may be difficult to accurately deduce the reactive load from the active load and vice versa. Another natural idea is to directly extend the methods designed for active load forecasting to reactive load forecasting; however, it is difficult to achieve ideal results in this way because of the different characteristics of the two kinds of loads. The active load is easily affected by external factors such as weather due to its close relationship with the user operation situation. Accordingly, the methodologies applied for active load forecasting mostly rely on analyzing the past loads in combination with multiple contemporaneous external factors to predict the future load through statistical analysis or artificial intelligence [9]. The response of reactive power to external factors is not as sensitive as that of active power. In addition, reactive power itself is local and decentralized, and its power flow direction is not as clear as that of active power.

Motivated by the fact that short-term forecasting can be divided into two strongly correlated subtasks, multi-task learning (MTL) is considered an effective way to approach this problem. In MTL, representation information can be shared among related subtasks to allow the original subtasks to be more effectively addressed [10]. MTL has been successfully applied in natural language processing [11], speech recognition [12], computer vision [13] and other fields. In this study, we propose a short-term power load forecasting method based on multi-task regression to utilize the intrinsic nonlinear relationship between active and reactive power, which can significantly improve the performance for each individual forecasting task.

Our contributions are listed as follows.

1. We provide a promising solution for short-term load forecasting. To the best of our knowledge, this is the first work on substation node reactive load forecasting via deep learning. The active and reactive loads are fed into an MTL model to perform joint forecasting so that the relationship between the two loads can provide constraints during the learning process to improve the forecasting accuracy.
2. We employ an attention mechanism to prevent performance deterioration of a long short-term memory (LSTM) network. In addition, the attention mechanism enables the extraction of more discriminative features, which have a positive influence on the prediction results.
3. We propose an adaptive combined-task-wise loss function for use in optimizing our model.
4. Extensive experiments are reported based on large amounts of real-world data collected from substations in different locations with different external situations. The results suggest that our proposed model is effective and robust, showing great potential for practical application.

The rest of the paper is organized as follows. Section 2 introduces the related work, and Section 3 describes our proposed method in detail. Section 4 elaborates on the construction of the datasets and the data preprocessing procedure. Section 5 presents the validation results and comparisons with other state-of-the-art load forecasting methods. Section 6 summarizes the advantages and disadvantages of the proposed method as well as future research.

2. Related work

Recently, most of the methodologies applied for active load forecasting have been based on statistical analysis or artificial intelligence. The commonly used statistical methods include multiple regression analysis, autoregressive moving average models and semiparametric additive models [14–16]. In the last decade, state-of-the-art machine-learning-based load forecasting methods have emerged. In some of them, features are designed and extracted through feature engineering

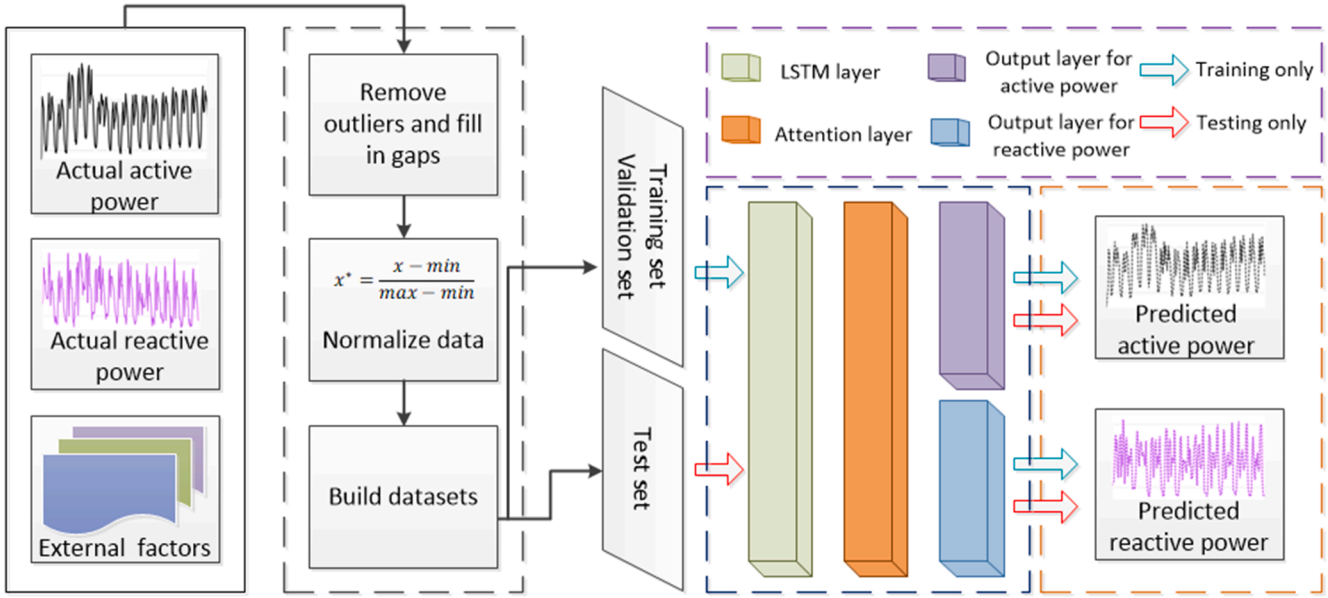


Fig. 2. Multitask load forecasting system.

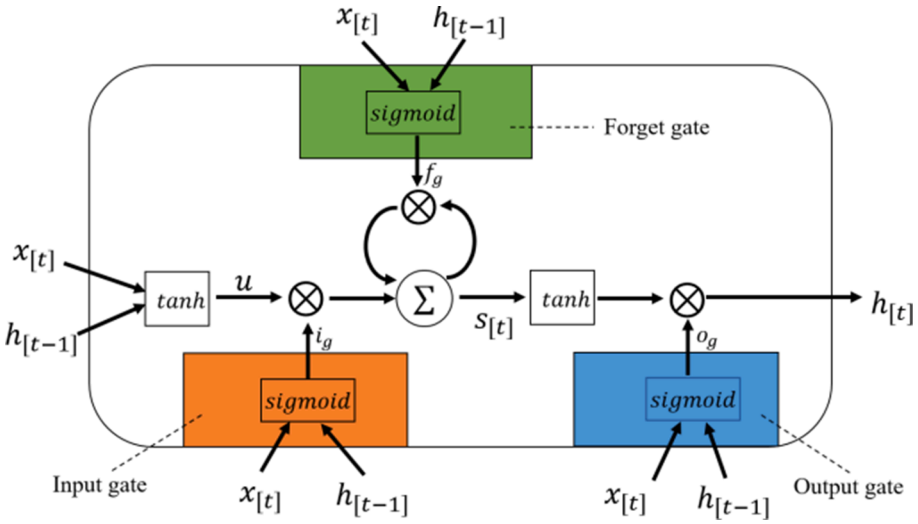


Fig. 3. The memory cell of LSTM. The orange box indicates the input gate, the green box indicates the forget gate and the blue box indicates the output gate.

methods and are then used for prediction based on traditional machine learning models such as support vector machines [17]. In others, advanced neural network architectures are employed to combine feature extraction and prediction in an end-to-end pipeline, as in LSTM-based or gated recurrent unit (GRU)-based methods [18–20]. Although reactive load forecasting is attracting increasing attention at present, successful achievements in this area are quite limited. Researchers from the University of Napoli Parthenope explored a short-term industrial reactive power forecasting method based on multiple linear regression and support vector regression (SVR) models [21]. Their validation showed that their model worked well at the levels of both aggregate and individual industrial loads. An artificial neural network (ANN)-based forecasting model for reactive loads has been utilized to solve the prediction problem under conditions of partial historical load data failure [22].

In recent decades, the LSTM architecture and its variants have become the most popular foundations for time sequence prediction and have begun to be widely used in various fields. In an electric power system, it is important to monitor the insulators in case they become disconnected, resulting in power failure. Failure identification can be performed through time series analysis based on audible signals generated from ultrasound, and several methods, including the LSTM approach and the wavelet-based group method of data handling (Wavelet GMDH), have been employed for this purpose [23]. Accurate short-term solar irradiance forecasting is critical to ensure the optimal utilization of photovoltaic power generation sources. A convolutional neural network (CNN)-LSTM model has been proposed for obtaining the relevant spatiotemporal correlations and predicting global horizontal radiance one hour in advance [24]. Fog forecasting is also important in many fields, including agriculture, economics, public health, and transportation. A novel LSTM framework composed of an LSTM network and a fully connected layer has been proposed for short-term fog forecasting [25]. In recent years, the installed capacity of renewable energy systems such as wind energy systems has increased significantly worldwide. Accurate wind speed prediction is therefore of significant importance for the development of the renewable energy market, and it has been shown that such prediction can be achieved by means of amplitude and frequency modulation (AM-FM) theory combined with an ensemble of variational mode decomposition (VMD) and single spectrum analysis (SSA)-LSTM models [26].

3. Proposed work

The pipeline of the proposed multi-task load forecasting strategy is shown in Fig. 2. There are three steps in the workflow. First, the historical loads and potential corresponding external factors are collected to construct the input data sequence. Second, the raw dataset is preprocessed to screen and remove outliers as well as to fill in missing values. Third, the data are fed into a multi-task load forecasting model to predict the reactive and active loads. The details of the model are described in the following subsections.

3.1. LSTM

The LSTM architecture is chosen as the backbone of our forecasting model since it can generate the future load based on the historical load sequence with high accuracy. The functional unit of an LSTM network is the memory cell shown in Fig. 3, through which the input sequence is propagated following the bold arrows. The input vector is denoted as $\boldsymbol{x}_{[t]}$, and the output vector of the LSTM unit is denoted as $\boldsymbol{h}_{[t]}$, where subscript $\boldsymbol{t}$ indicates the $\boldsymbol{t}^{\boldsymbol{th}}$ step in the sequence. There are three processing gates located in the LSTM unit, namely, an input gate, a forget gate and an output gate, as shown in the orange, the green and the blue boxes in Fig. 3 [27]. Let $\boldsymbol{i_g}, \boldsymbol{f_g}$ and $\boldsymbol{o_g}$ be the weights produced by the input gate, forget gate and output gate, respectively. During the $\boldsymbol{t}^{\boldsymbol{th}}$ step processing of the LSTM unit, the $\boldsymbol{x}_{[t]}$ and $\boldsymbol{h}_{[t-1]}$ are fed into the three gates directly to produce corresponding weights. Meanwhile, they are employed to generate an updated vector $\boldsymbol{u}$. The working protocol is described in Eq. (1) to Eq. (4), where $\boldsymbol{W..}$ and $\boldsymbol{b..}$ denote a trainable linear transform and a bias, respectively.

$$i_g = sigmoid\left(x_{[t]}W_{ix} + h_{[t-1]}W_{ih} + b_i\right) \tag{1}$$

$$f_g = sigmoid\left(x_{[t]}W_{fx} + h_{[t-1]}W_{fh} + b_f\right) \tag{2}$$

$$o_g = sigmoid\left(x_{[t]}W_{ox} + h_{[t-1]}W_{oh} + b_o\right) \tag{3}$$

$$u = tanh\left(x_{[t]}W_{ux} + h_{[t-1]}W_{uh} + b_u\right) \tag{4}$$

During propagation of the input $x_{[t]}$ in the LSTM unit, the obtained three nonlinear weights affect the updated vector u through matrix multiplication operations. Following the bold arrows, the updated vector u is first weighted by the input gate weight i_g and is then summed with the forget-gate-weighted $s_{[t-1]}$ to obtain the new $s_{[t]}$, which is called the cell state in the t^{th} step. Finally, $s_{[t]}$ is nonlinearly activated by the *tanh* function and weighted by o_g to obtain the output vector of the LSTM unit in the t^{th} step, $h_{[t]}$, as shown in Eq. (5) and Eq. (6).

$$s_{[t]} = f_g \circ s_{[t-1]} + i_g \circ u \tag{5}$$

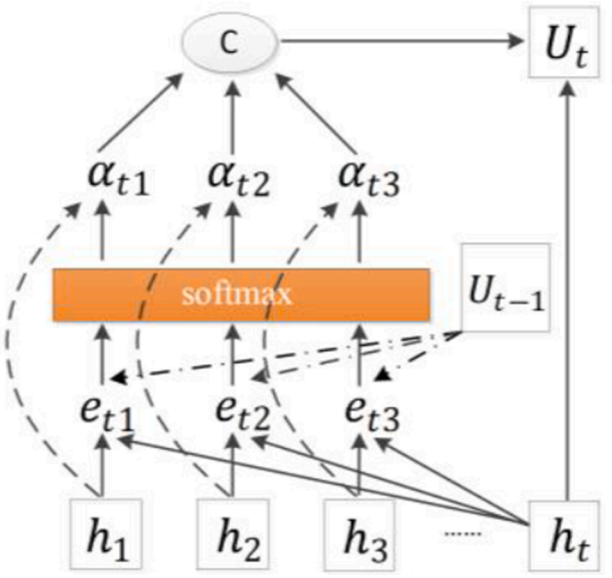


Fig. 4. Attention structure.

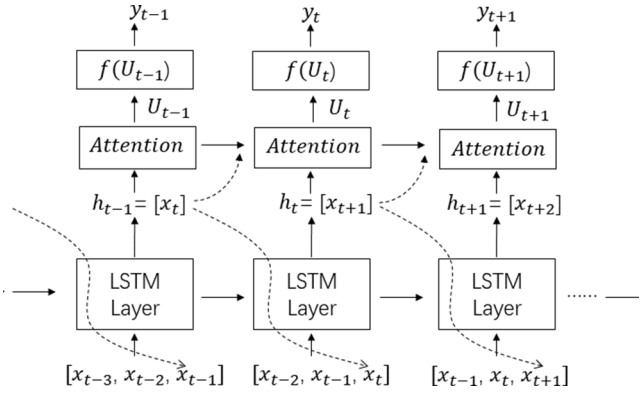


Fig. 5. Attention-LSTM network.

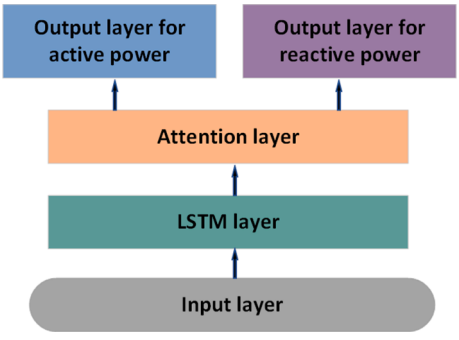


Fig. 6. Multi-task load forecasting model.

$$h_{[t]} = o_g{}^{\circ}\tanh\left(s_{[t]}\right) \tag{6}$$

3.2. Attention mechanism in the LSTM model

An attention mechanism [28,29] is added to the backbone LSTM model to enforce its ability to focus on critical features from the long input sequence. The principle of the attention mechanism used in this paper is illustrated in Fig. 4. The input sequence $\{h_1, h_2, h_3, \cdots, h_t\}$ consists of the outputs of the LSTM layer in different time steps. We calculate the attention coefficient C, which represents the contribution of the previous sequence $\{h_1, h_2, h_3, \cdots, h_{t-1}\}$ to the current h_t. e_{ti}, as calculated via Eq. (7), measures the relationship between the output vectors in the t^{th} time step and the current t^{th} step. The softmax function is used to convert e_{ti} into a probability-like parameter α_{ti}, as shown in Eq. (8).

$$e_{ti} = U_{t-1}\tanh(V_{t0}h_t + V_{ti}h_i + b_{ti}) \tag{7}$$

$$\alpha_{ti} = softmax(e_{ti}) = \frac{\exp(e_{ti})}{\sum_j^N \exp(e_{tj})} \tag{8}$$

where U_{t-1} is the hidden-state attention value in time step $t-1$. Then, the weighted sum of the α_{ti} and h_i is calculated to obtain the attention coefficient C. In turn, h_t is weighted by the attention coefficient C to obtain U_t as follows:

$$C = \sum_i^N \alpha_{ti} h_i \tag{9}$$

$$U_t = C^* h_t \tag{10}$$

Our base attention-LSTM load forecasting model, as shown in Fig. 5, calculates the final predicted load in the t^{th} step by applying Eq. (11). For a general single-task prediction model, y_t refers to either the reactive load or the active load. We adopt MTL to allow the model to output the predicted reactive and active loads simultaneously, as described in the next section.

$$y_t = sigmoid(U_t) \tag{11}$$

3.3. Multi-task attention-LSTM model

To perform accurate short-term forecasting of the active and reactive loads simultaneously, we build a multi-task regression model with hard parameter sharing based on the basic attention-LSTM model introduced above. The two subtasks, namely, active load forecasting and reactive load forecasting, share the LSTM layer and the attention layer. Two separate output layers are then used to output the active and reactive power prediction results. The multi-task attention-LSTM (MTAL) model is illustrated in Fig. 6. Here we employ a one-layer LSTM model with 64 cells in it.

In MTL with hard parameter sharing, the bias among the subtasks is adjusted by means of a loss function. We propose an adaptive combined-task-wise loss function for use in optimizing our MTAL model, as shown in Eq. (12). The combined loss consists of two mean squared error (MSE) losses [30], one for each subtask, to balance the two tasks during training. In this way, the prediction accuracy can be better than that of a single-task model for both reactive and active load forecasting.

$$Loss = \gamma^* \frac{1}{n}\sum_{i=1}^{n}(p_i - \widehat{p_i})^2 + \lambda^* \frac{1}{n}\sum_{i=1}^{n}(q_i - \widehat{q_i})^2 \tag{12}$$

where p_i and q_i represent the actual active and reactive loads, respectively; $\widehat{p_i}$ and $\widehat{q_i}$ are the predicted active and reactive loads, respectively; and γ and λ are hyperparameters used to balance the final accuracy. Through proper selection of these hyperparameters, overfitting and inclination toward one subtask over the other can be avoided, and the efficiency of our model can be improved. We discuss the hyperparameter selection process in the next subsection.

3.4. Grid search algorithm for parameter setting in the combined-task-wise loss function

We use k-fold cross-validation [31] and a grid search algorithm [32] to adaptively optimize the two hyperparameters, γ and λ. Grid search is a parameter optimization technique that is often used to optimize three or fewer hyperparameters. First, a multidimensional grid is formed in

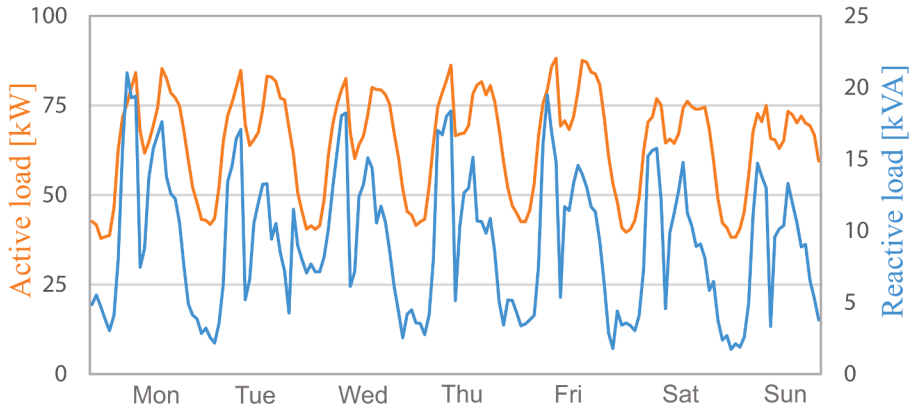


Fig. 7. Aggregate active and reactive power profiles (December 3nd-9th, 2018).

Table 1
Summary of external influencing factors.

	Factor name	Factor number
Weather factor	Maximum temperature/°C	1
	Minimum temperature/°C	2
	Average temperature/°C	3
	Weather type	4
	Wind speed/m/s	5
	Humidity/%	6
Calendar factor	Weekday or not	7
	Holiday or not	8
	Time/min	9

accordance with the value ranges of the parameters. Each node in the grid represents a combination of parameter values. Second, the weights in the loss function are determined in accordance with each grid node to train the MTAL model. Each group of grid nodes corresponds to a group of hyperparameters. Third, *k*-fold cross-validation is used to evaluate the selected parameters to reliably find the optimal parameter combination. The root mean square errors (RMSEs) for the active and reactive load forecasting subtasks in the j^{th} test set are calculated and recorded as r_{pj} and r_{qj}, respectively. Then, the mean of all r_{pj} and r_{qj}, denoted by R, is derived as the evaluation metric for the model corresponding to the current grid node, as follows:

$$r_p = \sqrt{\frac{\sum_{i=1}^{n}(p_i - \widehat{p}_i)^2}{n}} \tag{13}$$

$$r_q = \sqrt{\frac{\sum_{i=1}^{n}(q_i - \widehat{q}_i)^2}{n}} \tag{14}$$

$$R = \frac{1}{2} * \left[\frac{\sum_{j=1}^{k}(r_{pj})}{k} + \frac{\sum_{j=1}^{k}(r_{qj})}{k}\right] \tag{15}$$

To improve the accuracy of parameter optimization, we use a two-stage coarse-to-fine search strategy in our grid search process: the first stage is a coarse search in steps of 0.1, and the second stage is a fine search in steps of 0.01. Finally, the best hyperparameters γ and λ are determined to be those that yield the minimum value of R.

4. Datasets and experimental preparation

4.1. Datasets

Historical load: Five substations were chosen to validate the robustness of the proposed method. Two of these substations, ST-A and ST-B, are located in Northwest China, where the electricity structure is complex and new energy power accounts for a large proportion of the power supply. The remaining substations, ST-C, ST-D and ST-E, are located in a coastal city in Southeast China, where hydropower stations and coal-fired power stations are the major power suppliers. The load distributions of the chosen substations are very different due to difference in weather conditions, industrial development and local economic situations. We collected active and reactive power data every 15 min. The case study in this paper is based on data measured from July 1 to December 31, 2018. Fig. 7 shows the total active and reactive power profiles throughout week at ST-A (December 3rd-9th, 2018).

External factors: As mentioned before, the load is very sensitive to changes in external factors. Thus, in addition to historical active load and load data, the corresponding external factors should be considered. The external factors that have a direct influence on load forecasting can be mainly divided into two categories: calendar variables and weather conditions.

Calendar variables affect both customer behavior and production behavior, thereby affecting the load. For example, the amount of electricity used on weekdays is greater than that on weekends and holidays. Normally, production behavior is regular on daily and weekly scales. Therefore, the load follows similar patterns. Holidays are not as periodic, but they have a great influence on both customer and production behavior, and their effects on power loads cannot be neglected. Therefore, we use three variables to represent the calendar factors influencing each moment in time, namely, whether the current day is a weekday, whether the current day is a holiday, and the time of day.

The local weather conditions also affect power loads. We mainly consider the temperature, weather type, wind speed and humidity to describe the weather conditions. The chosen temperature features include the daily maximum, minimum and average temperatures. The possible weather types are categorized as sunny, rainy and snowy.

We construct a vector X_t to represent the data at time t, which can be expressed as

$$X_t = [l_{t1}, l_{t2}, \cdots, l_{t9}, p_t, q_t] \tag{16}$$

where $\{l_{t1}, l_{t2}, \cdots, l_{t9}\}$ are the external factors at time t, as shown in Table 1, and p_t and q_t are the actual active and reactive power levels, respectively, at time t.

4.2. Data preprocessing

The historical load data were acquired directly from the supervisory control and data acquisition (SCADA) system of the grid company. It is difficult to avoid abnormal or missing data due to potential failures of the SCADA system. To address this problem, we used the boxplot method to find abnormal data and removed them, and Lagrange interpolation [33] was employed to fill in missing values. Afterward, we normalized the original data to the range of [0,1] to unify their

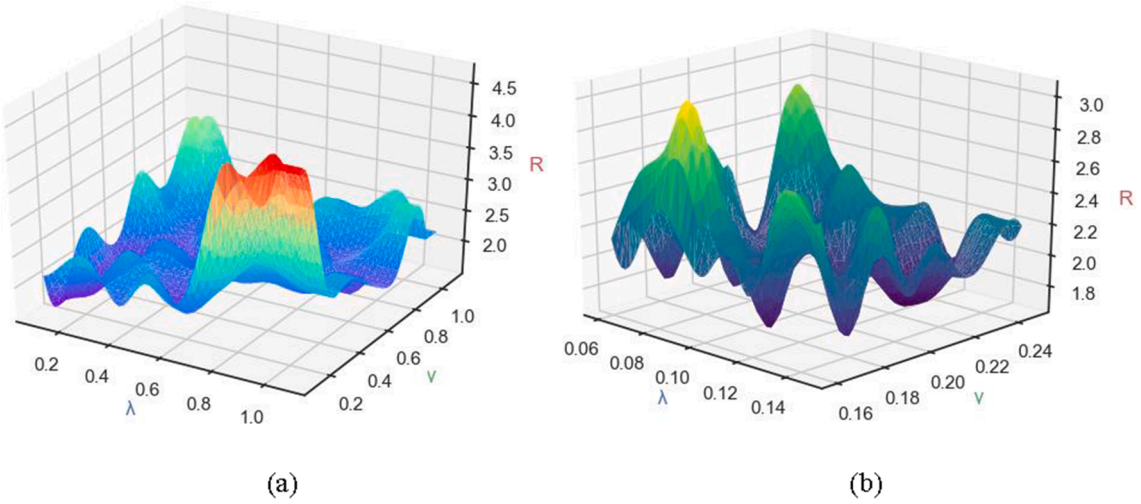


Fig. 8. (a) Results of the coarse grid search, with a step size of 0.1. (b) Results of the fine grid search, with a step size of 0.01.

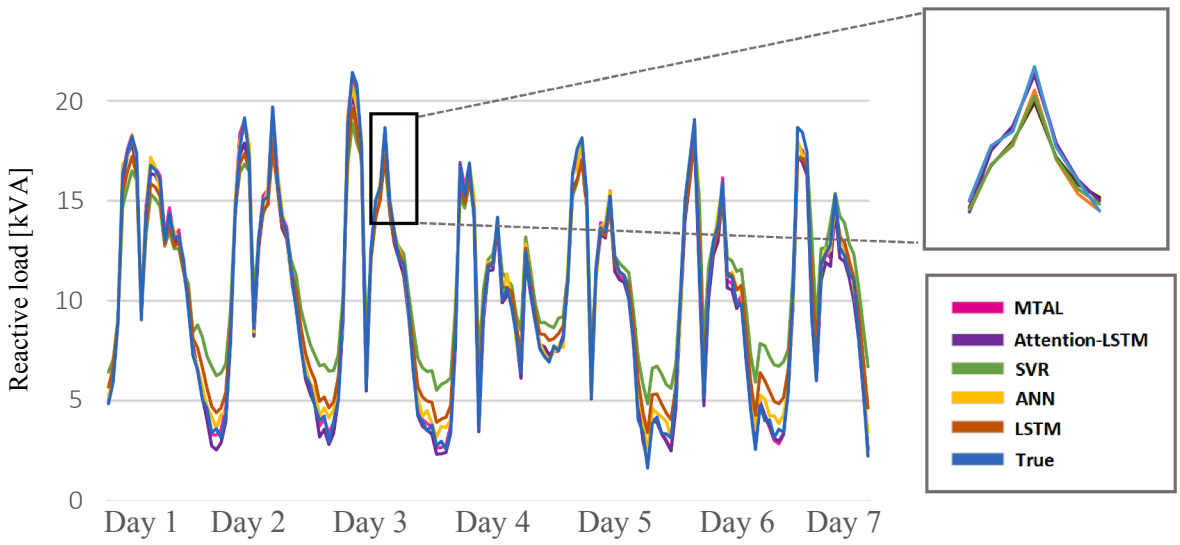


Fig. 9. Reactive load prediction curves for the one-week test set.

Table 2
RMSEs and MAPEs of the reactive load prediction models for substation ST-A.

	One-month test set		One-week test set		One-day test set	
	RMSE	MAPE/ %	RMSE	MAPE/ %	RMSE	MAPE/ %
MTAL	**0.552**	**4.364**	**0.597**	**2.935**	**0.554**	**1.922**
attention-LSTM	0.613	4.778	0.682	3.695	0.561	4.581
SVR	0.767	5.313	0.884	4.609	0.627	4.336
ANN	0.712	5.890	0.707	5.062	0.614	5.279
LSTM	0.781	5.771	0.833	4.953	0.601	5.322

Table 3
RMSEs and MAPEs of the active load prediction models for substation ST-A.

	One-month test set		One-week test set		One-day test set	
	RMSE	MAPE/ %	RMSE	MAPE/ %	RMSE	MAPE/ %
MTAL	**1.618**	**2.324**	**3.008**	**4.537**	**2.741**	**3.533**
attention-LSTM	1.735	2.340	5.398	8.624	2.760	3.721
LSTM-CNN	3.689	4.496	4.553	5.825	3.428	4.619
RNN	3.982	6.056	4.667	7.554	4.676	5.284
LSTM	2.358	4.163	5.278	8.357	3.081	4.132

dimensions. The formulas are as follows:

$$p^* = \frac{p - P_{min}}{P_{max} - P_{min}} \tag{17}$$

$$q^* = \frac{q - Q_{min}}{Q_{max} - Q_{min}} \tag{18}$$

where p^* and q^* are the normalized values of the active and reactive loads, respectively; P_{min} and Q_{min} are the corresponding minimum values; and P_{max} and Q_{max} are the corresponding maximum values.

We divided the data in chronological order. The data from July to November were used to construct the training set. Notably, the prediction accuracy of a model over different time horizons can reflect its robustness. Therefore, we used the data from December 1 as a one-day test set, the data from the first week of December as a one-week test set, and all data from December as a one-month test set. We calculate the forecasting accuracy on the three sets respectively to prove that our model still works well as time goes by.

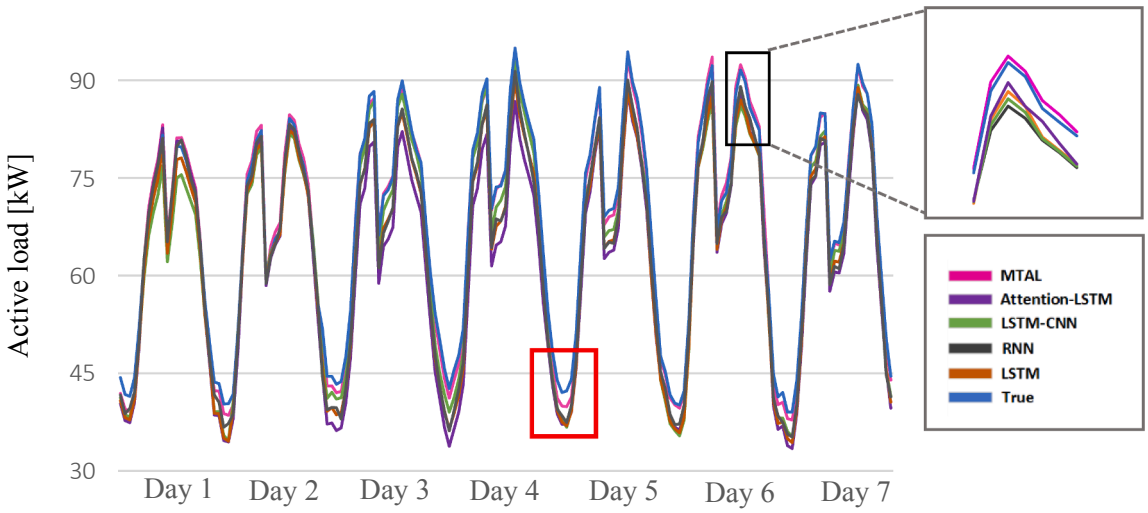


Fig. 10. Active load prediction curves for the one-week test set.

Table 4
Hyperparameters of the MTAL models for substations ST-B, ST-C, ST-D and ST-E.

	ST-B	ST-C	ST-D	ST-E
γ	0.21	0.16	0.25	0.22
λ	0.26	0.25	0.19	0.23

4.3. Evaluation metrics

In this paper, we choose the RMSE and the mean absolute percentage error (MAPE) as error evaluation metrics. The formulas are given as follows:

$$e_{p-MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{p_i - \widehat{p}_i}{p_i}\right| \times 100\% \tag{19}$$

$$e_{q-MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{q_i - \widehat{q}_i}{q_i}\right| \times 100\% \tag{20}$$

$$e_{p-RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(p_i - \widehat{p}_i)^2}{n}} \tag{21}$$

$$e_{q-RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(q_i - \widehat{q}_i)^2}{n}} \tag{22}$$

5. Results

All experiments were conducted on a PC equipped with an Intel® i5 CPU, 16 GB of RAM, and an NVIDIA® GTX 1080 Ti GPU. We first present the chosen hyperparameters and the results for ST-A in detail, and we then briefly report the results for the remaining four substations ST-B, ST-C, ST-D and ST-E in Section 5.3.

5.1. Grid search for the values of the hyperparameters in the loss function

We used 5-fold cross-validation and a grid search algorithm to determine the optimal hyperparameter values for the combined-task-wise loss function. The hyperparameters γ and λ were allowed vary in the range [0.1, 1]. We traversed all grid nodes and trained the model

Table 5
RMSEs and MAPEs of the active load prediction models for substations. T-B, ST-C, ST-D and ST-E.

		One month		One week		One day	
		RMSE	MAPE/%	RMSE	MAPE/%	RMSE	MAPE/%
ST-B	MTAL	**4.497**	**3.746**	**2.723**	**2.991**	**2.823**	**3.013**
	attention-LSTM	4.913	3.88	2.799	3.613	2.837	3.162
	LSTM-CNN	6.990	4.994	3.121	3.912	2.991	3.289
	RNN	4.570	4.447	2.904	3.904	2.840	3.295
	LSTM	5.107	4.255	2.823	4.738	3.220	4.175
ST-C	MTAL	**0.492**	**4.334**	**0.533**	**5.720**	**0.640**	**5.877**
	attention-LSTM	0.521	4.924	0.543	6.119	0.653	6.165
	LSTM-CNN	0.582	6.341	0.571	6.785	0.669	6.247
	RNN	0.538	5.200	0.566	6.905	0.647	6.32
	LSTM	0.632	6.247	0.589	7.084	0.672	6.723
ST-D	MTAL	**2.995**	**5.001**	**2.691**	**5.036**	**1.991**	**4.748**
	attention-LSTM	5.159	7.351	4.373	7.008	3.527	6.091
	LSTM-CNN	8.554	15.859	8.303	16.743	6.246	19.371
	RNN	9.552	18.421	8.622	19.633	8.31	25.666
	LSTM	7.318	8.339	5.823	6.031	3.875	5.645
ST-E	MTAL	**3.164**	**5.267**	**3.579**	**6.610**	**3.265**	**3.543**
	attention-LSTM	4.579	5.935	4.031	6.882	4.559	3.963
	LSTM-CNN	5.881	9.823	4.258	7.310	4.826	6.327
	RNN	5.273	10.220	5.053	7.797	5.826	5.227
	LSTM	4.984	8.375	6.228	7.022	8.652	9.254

Table 6
RMSEs and MAPEs of the reactive load prediction models for substations. ST-B, ST-C, ST-D and ST-E.

		One month		One week		One day	
		RMSE	MAPE/%	RMSE	MAPE/%	RMSE	MAPE/%
ST-B	MTAL	**2.336**	**5.725**	**2.155**	**3.968**	**3.174**	**3.401**
	attention-LSTM	2.937	9.524	2.246	9.967	3.614	17.165
	ANN	3.772	13.948	2.806	12.764	4.333	20.826
	SVR	5.857	13.61	2.616	11.689	4.027	18.331
	LSTM	3.826	10.003	3.219	11.021	4.825	10.544
ST-C	MTAL	**0.491**	**4.359**	**0.532**	**5.706**	**0.641**	**5.88**
	attention-LSTM	0.517	4.969	0.542	6.137	0.656	6.222
	ANN	0.525	5.143	0.625	6.561	0.704	6.436
	SVR	0.689	5.607	0.682	7.395	0.752	6.885
	LSTM	0.626	5.208	0.657	7.053	0.718	6.019
ST-D	MTAL	2.395	**4.503**	**2.275**	**5.101**	**1.601**	**6.028**
	attention-LSTM	2.582	6.595	2.583	5.297	1.854	9.22
	ANN	2.79	17.655	3.433	26.49	3.053	14.232
	SVR	**2.045**	15.897	2.509	14.267	2.362	23.991
	LSTM	2.654	8.629	3.652	7.038	2.660	3.672
ST-E	MTAL	**3.274**	**5.829**	**3.527**	**5.281**	**2.476**	**6.261**
	attention-LSTM	4.628	5.925	4.279	5.923	5.241	8.563
	ANN	7.638	10.532	5.826	7.125	6.252	8.428
	SVR	10.632	7.628	7.293	10.532	11.532	13.253
	LSTM	6.632	8.521	4.620	8.628	9.252	10.742

using the loss function defined with the hyperparameter values corresponding to each node. For each trained model, we calculated the R value of the reactive and active power forecasting performance according to Eq. (15). Fig. 8 shows the results of hyperparameter optimization via grid search.

As seen from Fig. 8(a), the coarse search stage revealed that when λ is 0.1 and γ is 0.2, R achieves the minimum value of 0.786. In addition, we found that as λ decreases and γ increases, the value of R increases sharply, indicating that the prediction performance decreases rapidly. Next, we set the λ interval to [0.06,0.15] and the γ interval to [0.16,0.25] and performed the fine search stage, with a step size of 0.01. The experimental results are shown in Fig. 8(b). It can be seen from this figure that when λ is 0.12 and γ is 0.23, the minimum R value of 0.859 is achieved.

5.2. Performance of the proposed MTAL model

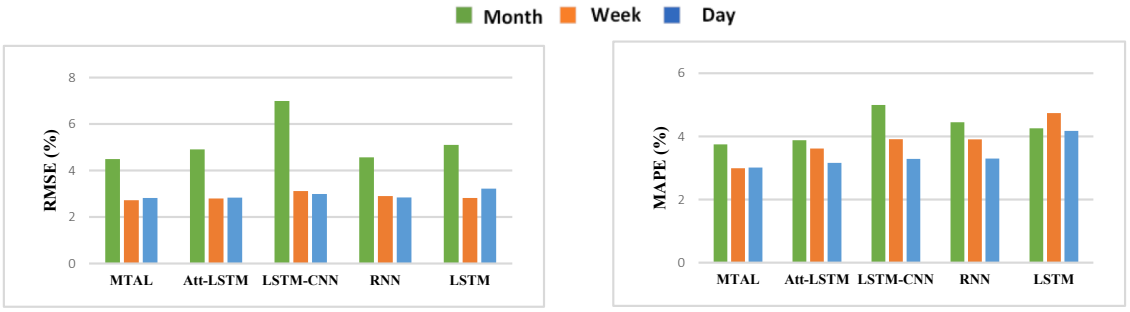
We validated the performance of our MTAL model from two perspectives. On the one hand, we attempted to demonstrate the effectiveness of the proposed multi-task strategy. To achieve this goal, we first trained the attention-LSTM model in single-task mode. That is, the attention-LSTM model was trained separately for active load forecasting and reactive load forecasting. On the other hand, we compared our MTAL method with other popular forecasting models for both of our subtasks at different substations to validate its robustness. For fairness, all models were trained using the same dataset and settings. We used the relu activation function instead of the sigmoid activation function to avoid the vanishing gradient problem [34].

Reactive load forecasting task: To date, the reported results on reactive load forecasting methods have been quite limited because researchers have only recently started to focus on this field. To the best of our knowledge, there are only two studies in which reactive load forecasting has been directly performed, based on an SVR model and an ANN model. Therefore, we compared our method with both of these models. Fig. 9 shows the prediction curves of the different models on the one-week test set as well as a zoomed-in view of the region in the black box. The blue curve represents the actual reactive load, and it is best fit by the MTAL prediction curve (red). The SVR and LSTM results diverge from the actual load values almost every night, as their predictions at midnight are always too high. The ANN model is better than the SVR model but still misses the load valley at midnight. The performance of the attention-LSTM model trained on a single task is similar to that of MTAL but is poorer near the inflection points. In other words, this model responds more slowly to sudden changes in load than the multi-task-trained model does. We believe that this might be because the MTL strategy uses information from the other subtask to compensate for deficiencies in reactive load forecasting and to enforce its generalizability. The overall prediction accuracies of the models on the test sets with different time horizons are shown in Table 2. It is expected that the prediction error will accumulate as the forecasting horizon increases. However, as seen from Table 2, the performance of MTAL does not degrade as much as that of the rest of the models, indicating that the proposed model is more robust.

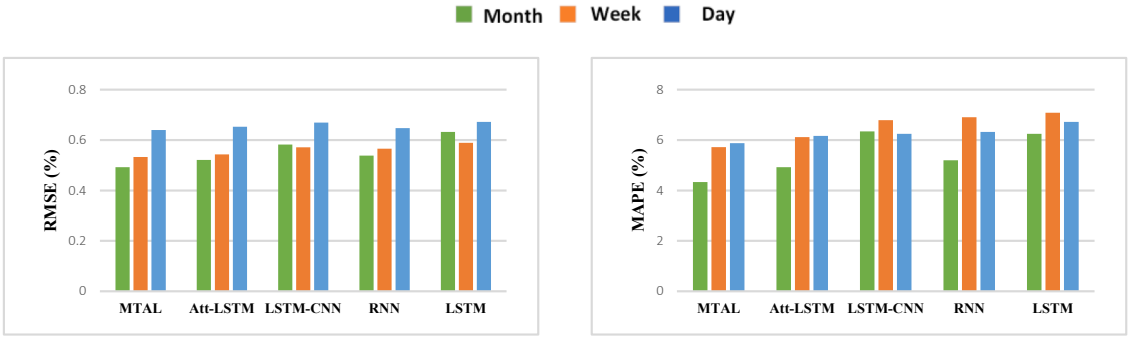
Active load forecasting task: Several state-of-the-art methods have focused on active load forecasting in recent years; among these method, sequence-to-sequence methods are the most popular. We chose two representative models, an LSTM-CNN model [35] and a recurrent neural network (RNN) model [36], as well as our single-task-trained attention-LSTM model for comparison with the proposed MTAL model. Table 3 shows the prediction accuracies of each model on the one-month, one-week and one-day test sets. Based on these results, the average prediction error of MTAL is the smallest. On the one-month test set, the RMSE value of MTAL is decreased by 0.117, 2.071, 2.364, and 0.74 compared with those of the attention-LSTM, LSTM-CNN, RNN and LSTM models, respectively, and the MAPE value is decreased by 0.016%, 2.172%, 3.732%, and 1.839%, respectively. On the one-week test set, the RMSE value of MTAL is decreased by 2.39, 1.545, 1.659, and 2.27 compared with those of the attention-LSTM, LSTM-CNN, RNN and LST models, respectively, and the MAPE value is decreased by 4.087%, 1.288%, 3.017%, and 3.82%, respectively. Fig. 10 illustrates the active load forecasting results for the one-week test set. Similar to the case of the reactive load forecasting subtask, the MTAL model performs extremely well at sharp peaks and valleys. These findings illustrate that the physical constraints between the two subtasks are indeed learned by the MTAL model and help to improve the forecasting performance.

5.3. Experimental analysis of substations in northwest China

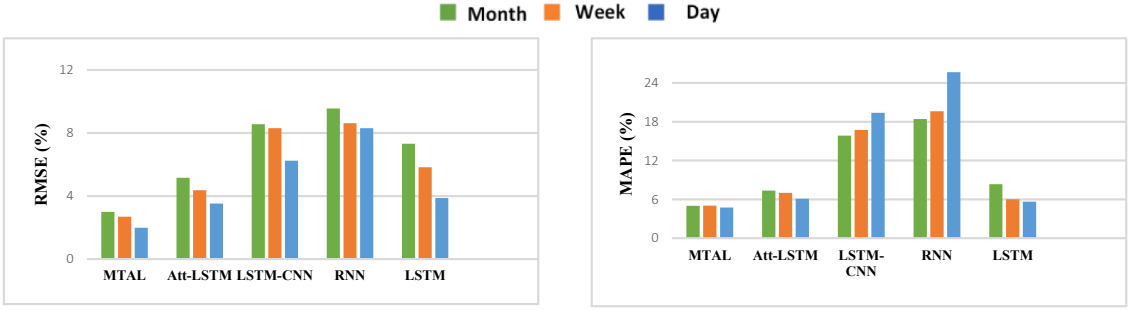
To further validate the effectiveness and robustness of our MTAL model, we repeated the experiments presented above for substations ST-



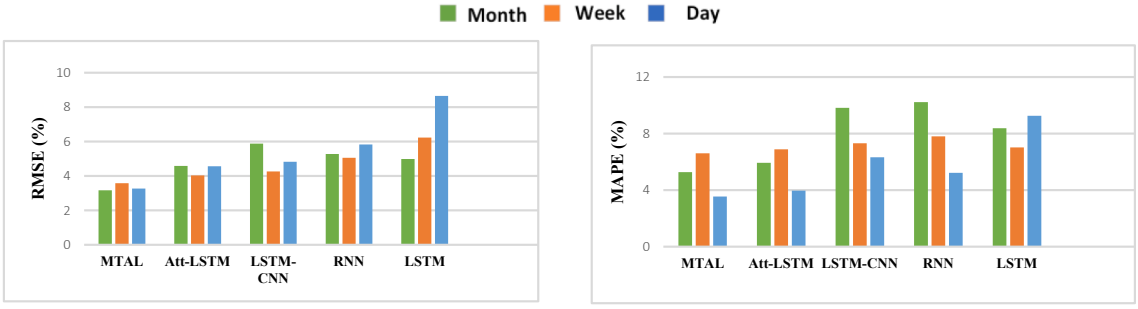
(a) Substation ST-B



(b) Substation ST-C



(c) Substation ST-D



(d) Substation ST-E

Fig. 11. Evaluation metric values for active load forecasting obtained by the different models at the four substations.

B, ST-C, ST-D and ST-E. The data were processed following the same protocol described previously. To obtain the hyperparameters γ and λ of the MTAL model corresponding to each station, we carried out the same grid search process based on the training data for each station. The hyperparameter results are shown in Table 4, and the final experimental results are shown in Table 5 and Table 6. It can be concluded from these tables that the proposed MTAL model performs robustly for any kind of power supply structure and consumer behavior. The evaluation metric values for active and reactive load forecasting obtained by the different models at the four substations are shown in Fig. 11 and Fig. 12 respectively.

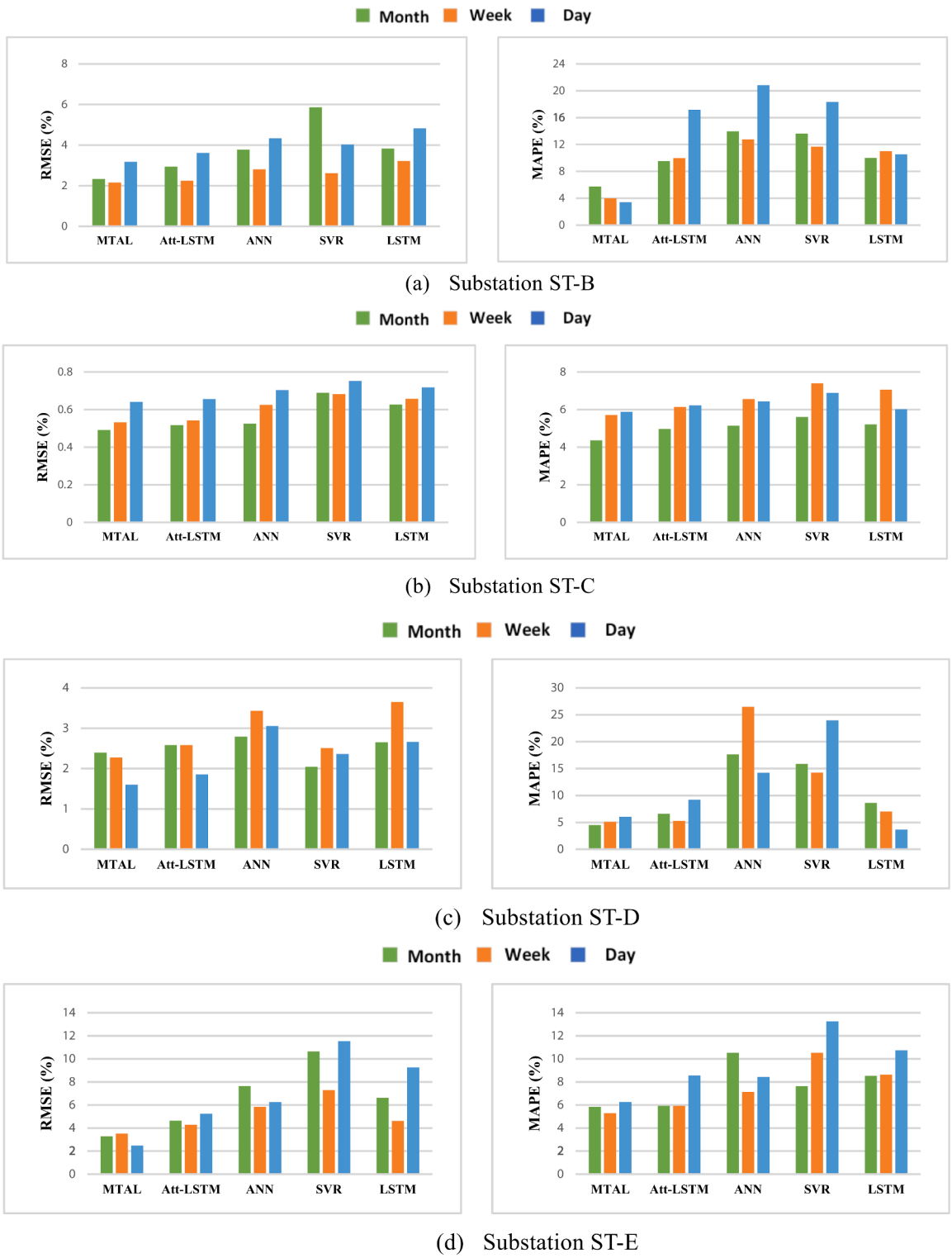


Fig. 12. Evaluation metric values for reactive load forecasting obtained by the different models at the four substations.

5.4. Discussion

In this paper, we have designed a load forecasting strategy for simultaneously predicting the active and reactive loads at substations. The proposed multi-task attention-LSTM model is abbreviated as MTAL. The main components of the architecture are a cascaded LSTM layer and an attention layer. When trained using the multi-task approach, the MTAL model can achieve good performance due to its effective feature extraction and the shared settings for the two complementary tasks.

The MTAL model was compared with the corresponding single-task attention-LSTM models for both subtasks. The only difference between MTAL and a single-task attention-LSTM model is that the former is trained in multi-task mode, while the latter is trained on an individual subtask. As shown in Table 2, MTAL yields smaller RMSE and MAPE values than the single-task attention-LSTM model when applied for reactive load prediction at ST-A. Similar results can be seen for the active

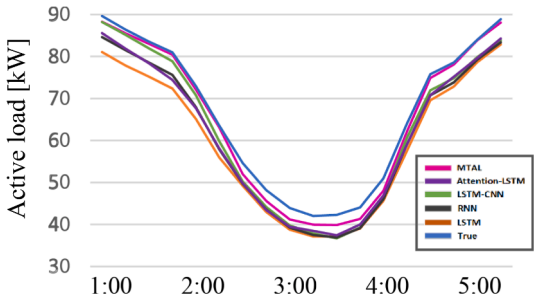


Fig. 13. Zoomed-in view of Fig. 10.

load prediction task in Table 3, and for the remaining four substations, the same results are also observed from Table 5 and Table 6. These findings demonstrate the effectiveness and necessity of MTL in MTAL. Due to the MTL strategy, MTAL can extract complementary features for active and reactive load forecasting and make full use of the hidden constraints in the physical power system to improve the forecasting accuracy.

We insert an attention layer following the LSTM layer in the MTAL model. The aim is to mitigate the performance degradation of a pure LSTM network when confronted with a long input sequence. Comparative experiments were performed between MTAL and the original LSTM model to validate the effectiveness of the proposed architecture. From Table 2 to Table 6, it is obvious that when the attention-LSTM model is trained in the single-task mode, it achieves better RMSE and MAPE values than the pure LSTM model. When MTL is further applied in MTAL, the performance is further improved. These findings demonstrate that the combination of an attention mechanism and the LSTM architecture can improve the ability of the model to handle long input sequences.

The data used in this paper were collected from five power substations located in the northwest and southeast coastal areas of China. Since the environments of the five substations are completely distinct from each other, differing, for example, in their energy types and levels of local economic development, we believe that the results are sufficient to validate the robustness of our method. In addition, to the best of our knowledge, there is no public dataset currently available for reactive load forecasting, and our dataset might be helpful for future research in this field. Based on this dataset, we performed load forecasting using MTAL for one-month, one-week and one-day time horizons to fully explore the performance of our model at different prediction scales. Comparisons of MTAL with other popular load forecasting models for both reactive and active loads were conducted. We adopted SVR and ANN models for reactive load forecasting since they are the only previous methods specifically designed for reactive load forecasting. As shown in Table 2, they perform worse than MTAL on the ST-A dataset. For active load forecasting, we chose LSTM-CNN and RNN models for comparison because of the representative nature of these two methods. As shown in Table 3, MTAL still achieves the best RMSEs and MAPEs among all considered methods on the ST-A dataset. Similar results can be observed from the forecasting results for the rest of the substations shown in Table 5 and Table 6. In other words, our MTAL is an accurate forecasting model that outperforms other popular load forecasting methods. In addition, it performs robustly and stably.

Although impressive performance has been achieved, there are still two limitations that need to be addressed in our future work. First, the proposed model is somewhat time consuming due to the large number of model parameters; second, the performance at points of sudden change in the load curve still needs to be improved. In Fig. 13, we present a zoomed-in view of the region indicated by the red box in Fig. 10 to visualize the results more closely. The prediction performance of our MTAL model is better than that of the other methods, and the predicted curve of MTAL is closer to the shape of the real load curve. However, there are still small differences between the predicted load values and the real load values near the inflection point, and a small gap can be observed between the curves. We speculate that this is because the load changes too sharply in the middle of the night. One potential solution is to explore the local load structure more thoroughly and to target corresponding external factors based on the specific energy composition of the power that is produced and consumed.

6. Conclusion and future research

In this paper, we have proposed a multi-task attention-LSTM (MTAL) model based on hard parameter sharing for reactive and active load forecasting. We have designed a combined-task-wise loss function for training the proposed model to perform multi-task load forecasting. The hyperparameters of the loss function are adaptively optimized through cross-validation and a grid search strategy. In addition to the historical active and reactive load data, we consider the influence of external factors, namely, calendar variables and weather conditions, to expand the data dimensions and improve the forecasting accuracy. For both subtasks, i.e., active load forecasting and reactive load forecasting, experimental results show that our MTAL model is generally superior to other methods on various test sets, indicating that our model offers good prediction performance and robustness and could be a potential solution for practical application in power systems. In future research, we will prune the model to reduce its complexity and focus on accelerating the model training and inference processes. In addition, we will further refine the parameters for load forecasting at peaks and valleys to improve the overall forecasting performance of the model.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijepes.2021.107517.

References

[1] Ali M, Adnan M, Tariq M. Optimum control strategies for short term load forecasting in smart grids. Int J Electr Power Energy Syst 2019;113:792–806. https://doi.org/10.1016/j.ijepes.2019.06.010.
[2] Darling GR, Wang ZS, Holloway S. Energy exchange in reactive scattering of hydrogen molecules from a Cu surface. Chem Phys Lett 2002;365(1-2):157–63. https://doi.org/10.1016/S0009-2614(02)01410-0.
[3] Jin H, Li Z, Sun H, Guo Q, Wang B. A two-stage reactive power optimization in transmission network incorporating reserves from voltage-dependent loads. Energy 2018;157:752–63. https://doi.org/10.1016/j.energy.2018.05.112.
[4] Byerly RT, Poznaniak DT, Taylor ER. Static Reactive Compensation for Power Transmission Systems. IEEE Power Eng Rev 1982;PER-2:52. DOI: 10.1109/MPER.1982.5519916.
[5] Mohamed SA. Enhancement of power quality for load compensation using three different FACTS devices based on optimized technique. Int Trans Electr Energy Syst 2020;30:1–25. https://doi.org/10.1002/2050-7038.12196.
[6] Wan C, Zhao J, Song Y, Xu Z, Lin J, Hu Z. Photovoltaic and solar power forecasting for smart grid energy management. CSEE J Power Energy Syst 2015;1(4):38–46. https://doi.org/10.17775/CSEEJPES.2015.00046.
[7] Rather ZH, Chen Z, Thogersen P, Lund P. Dynamic Reactive Power Compensation of Large-Scale Wind Integrated Power System. IEEE Trans Power Syst 2015;30(5):2516–26. https://doi.org/10.1109/TPWRS.2014.2365632.
[8] Zhong J, Nobile E, Bose A, Bhattacharya K. Localized reactive power markets using the concept of voltage control areas. IEEE Trans Power Syst 2004;19(3):1555–61. https://doi.org/10.1109/TPWRS.2004.831656.
[9] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. Int J Forecast 2016;32(3):914–38. https://doi.org/10.1016/j.ijforecast.2015.11.011.
[10] Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage 2012;59(2):895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069.

[11] Worsham J, Kalita J. Multi-task learning for natural language processing in the 2020s: Where are we going? Pattern Recognit Lett 2020;136:120–6. https://doi.org/10.1016/j.patrec.2020.05.031.
[12] Gong C, Li G, Luo L. Research on Acoustic Model of Multi-Task Learning for Speech Recognition. J Phys Conf Ser 2020;1550:032159. https://doi.org/10.1088/1742-6596/1550/3/032159.
[13] Huang H, Zhou H, Yang X, Zhang L, Qi L, Zang AY. Faster R-CNN for marine organisms detection and recognition using data augmentation. Neurocomputing 2019;337:372–84. https://doi.org/10.1016/j.neucom.2019.01.084.
[14] Roediger HL, Watson JM, McDermott KB, Gallo DA. Factors that determine false recall: A multiple regression analysis. Psychon Bull Rev 2001;8(3):385–407. https://doi.org/10.3758/BF03196177.
[15] Bayer FM, Bayer DM, Pumi G. Kumaraswamy autoregressive moving average models for double bounded environmental data. J Hydrol 2017;555:385–96. https://doi.org/10.1016/j.jhydrol.2017.10.006.
[16] Fan S, Hyndman RJ. Short-term load forecasting based on a semi-parametric additive model. IEEE Trans Power Syst 2012;27(1):134–41. https://doi.org/10.1109/TPWRS.2011.2162082.
[17] Zhang X, Wang J, Zhang K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. Electr Power Syst Res 2017;146:270–85. https://doi.org/10.1016/j.epsr.2017.01.035.
[18] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl Energy 2019;236:1078–88. https://doi.org/10.1016/j.apenergy.2018.12.042.
[19] Wang S, Wang X, Wang S, Wang D. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. Int J Electr Power Energy Syst 2019;109:470–9. https://doi.org/10.1016/j.ijepes.2019.02.022.
[20] Liu H, Mi X, Li Y, Duan Z, Xu Y. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression. Renew Energy 2019;143:842–54. https://doi.org/10.1016/j.renene.2019.05.039.
[21] Bracale A, Carpinelli G, De Falco P, Hong T. Short-term industrial reactive power forecasting. Int J Electr Power Energy Syst 2019;107:177–85. https://doi.org/10.1016/j.ijepes.2018.11.022.
[22] Fidalgo JN, Lopes JAP. Load forecasting performance enhancement when facing anomalous events. IEEE Trans Power Syst 2005;20:408–15. DOI: 10.1109/TPWRS.2004.840439.
[23] Stefenon SF, Dal Molin Ribeiro MH, Nied A, Mariani VC, Coelho LDS, Menegat da Rocha DF, et al. Wavelet group method of data handling for fault prediction in electrical power insulators. Int J Electr Power Energy Syst 2020;123:106269. https://doi.org/10.1016/j.ijepes.2020.106269.
[24] Zang H, Liu L, Sun L, Cheng L, Wei Z, Sun G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. Renew Energy 2020;160:26–41. https://doi.org/10.1016/j.renene.2020.05.150.
[25] Miao K-C, Han T-T, Yao Y-Q, Lu H, Chen P, Wang B, et al. Application of LSTM for short term fog forecasting based on meteorological elements. Neurocomputing 2020;408:285–91. https://doi.org/10.1016/j.neucom.2019.12.129.
[26] Rodrigues Moreno S, Gomes da Silva R, Cocco Mariani V, dos Santos Coelho L. Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network. Energy Convers Manag 2020;213:112869. https://doi.org/10.1016/j.enconman.2020.112869.
[27] Wang JQ, Du Yu, Wang J. LSTM based long-term energy consumption prediction with periodicity. Energy 2020;197:117197. https://doi.org/10.1016/j.energy.2020.117197.
[28] Choi M, Kim H, Han B, Xu N, Lee KM. Channel Attention Is All You Need for Video Frame Interpolation. Proc AAAI Conf Artif Intell 2020;34:10663–71. https://doi.org/10.1609/aaai.v34i07.6693.
[29] Li X, Zhang W, Ding Q. Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. Signal Process. 2019;161:136–54. https://doi.org/10.1016/j.sigpro.2019.03.019.
[30] Theodoridis S. Chapter 4 - Mean-Square Error Linear Estimation. In: Theodoridis SBT-ML, editor., Oxford: Academic Press; 2015, p. 105–60. DOI: 10.1016/B978-0-12-801522-3.00004-5.
[31] Barrow DK, Crone SF. Cross-validation aggregation for combining autoregressive neural network forecasts. Int J Forecast 2016;32(4):1120–37. https://doi.org/10.1016/j.ijforecast.2015.12.011.
[32] Lerman PM. Fitting Segmented Regression Models by Grid Search. Appl Stat 1980;29(1):77. https://doi.org/10.2307/2346413.
[33] Mastroianni G. Uniform convergence of derivatives of Lagrange interpolation. J Comput Appl Math 1992;43(1-2):37–51. https://doi.org/10.1016/0377-0427(92)90258-Y.
[34] Cui J, Qiu S, Jiang M, Pei Z, Lu Y. Text Classification Based on ReLU Activation Function of SAE Algorithm. In: Cong F., Leung A., Wei Q. (eds) Advances in Neural Networks - ISNN 2017. ISNN 2017. Lecture Notes in Computer Science, vol 10261. Springer, Cham. DOI: 10.1007/978-3-319-59072-1_6.
[35] Kim TY, Cho SB. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72–81. https://doi.org/10.1016/j.energy.2019.05.230.
[36] Pang Z, Niu F, O'Neill Z. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. Renew Energy 2020;156:279–89. https://doi.org/10.1016/j.renene.2020.04.042.