Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Power system load forecasting using mobility optimization and multi-task learning in COVID-19

Jiefeng Liu\textsuperscript{a}, Zhenhao Zhang\textsuperscript{a}, Xianhao Fan\textsuperscript{a}, Yiyi Zhang\textsuperscript{a,}\textsuperscript{*}, Jiaqi Wang\textsuperscript{a}, Ke Zhou\textsuperscript{b}, Shuo Liang\textsuperscript{b}, Xiaoyong Yu\textsuperscript{b}, Wei Zhang\textsuperscript{b}

\textsuperscript{a} Guangxi Key Laboratory of Power System Optimization and Energy Technology, Guangxi University, Nanning, Guangxi 530004, China
\textsuperscript{b} Guangxi Electric Power Research Institute, Nanning 530000, China

HIGHLIGHTS

\begin{itemize}
\item Considered the significant impact of the COVID-19 on electricity load.
\item Developed a novel method based on multi-task learning using parameters sharing layers.
\item Used the real mobility data from Google and Apple and load data from large power enterprises and public institutions.
\item The universally adaptable model can control the forecasting error at a low level in the COVID-19.
\item Implemented the Shapley additive explanations technology to enhance interpretability.
\end{itemize}

ARTICLE INFO

Keywords:
Short-term load forecasting
Long and short-term memory neural network
Multi-task learning
New coronavirus pandemic
Population mobility

ABSTRACT

Affected by the new coronavirus (COVID-19) pandemic, global energy production and consumption have changed a lot. It is unknown whether conventional short-term load forecasting methods based on single-task, single-region, and conventional indicators can accurately capture the load pattern during the COVID-19 and should be carefully studied. In this paper, we make the following contributions: 1) A mobility-optimized load forecasting method based on multi-task learning and long short-term memory network is innovatively proposed to alleviate the impact of the COVID-19 on short-term load forecasting. The incorporation of mobility data and data sharing layers potentially reduces the difficulty of capturing the load patterns and improves the generalization of the load forecasting models. 2) The real public data collected from multiple agencies and companies in the United States and European countries are used to conduct horizontal and vertical tests. These tests prove the failure of the conventional models and methods in the COVID-19 and demonstrate the high accuracy (error mostly less than 1%) and robustness of the proposed model. 3) The Shapley additive explanations technology based on game theory is innovatively introduced to improve the objectivity of the models. It visualizes that mobility indicators are of great help to the accurate load forecasting. Besides, the non-synchronous relationships between the indicators' correlations and contributions to the load have been proved.

1. Introduction

The new coronavirus 2019 (COVID-19) pandemic has affected almost all aspects of the society, and the power grid is no exception [1]. Unlike sudden extreme weather conditions (e.g. a snowstorm in Texas [2]) in certain areas and long-term stable social problems (e.g. ageing population [3]) in a country, the impact of the COVID-19 has spread all over the world, and the strength of the COVID-19 is deeply affected by human activities. During the outbreak of the COVID-19, many policies and measures have been implemented to curb the spread of the virus [4], such as blockade or partial blockade of cities, orders cancellation, curfews, shutdowns of industrial and mining enterprises, etc. At the same time, the power system’s load presents an entirely different operation pattern from normal working days and holidays [5]. Almost all traditional and even optimized load forecasting models seem powerless [6] when facing such public emergencies and crises.

Currently, the power system load forecasting can be divided into four types: ultra-short-term [7], short-term [8], medium-term [9], and long-
term [10]. Short-term load forecasting has been a research hotspot, and multi-feature load forecasting methods are commonly used. However, most existing studies only consider traditional features in the model establishment, such as load with time series, day type [11], weather [12], etc. It is unknown whether the models proposed in the previous studies will work adequately during COVID-19.

The existing load forecasting methods can be divided into conventional methods and machine learning methods. The conventional methods mainly include the time-series [13] analysis method and regression analysis [14]. Mature models like auto regression (AR) model, moving average (MA) model, or auto regressive moving average (ARIMA) model, are all load forecasting models based on stationary time series. However, the load changes with time are prone to non-stationary processes, so the accuracy of these models is not high. The auto regressive integrated moving average (ARIMA) [15] model is optimized for homogeneous non-stationary time series, but the problem of excessive dependence on historical load is still serious. Recently, the integrated energy system (IES) and machine learning methods have been developed. Machine learning methods have strong expressive power [16] and learning capabilities for non-linear data and time-series data [17], which have achieved impressive results in load forecasting. Some methods use swarm intelligence optimization algorithms to optimize support vector machines (SVM) and other machine learning processes, such as using grey wolf optimizer (GWO), firefly algorithm (FA), and particle swarm optimization (PSO) to construct GWO-SVM [18], FA-SVM [19], and PSO-SVM [20]. However, the mean absolute percentage Error (MAPE) values in the peak point of these studies are relatively high, and the auto-correlation problem in the load data is unsolved [21]. The traditional method to suppress the data auto-correlation is based on empirical mode decomposition (EMD) [22], but it has modal aliasing. To solve this problem, variational modal decomposition (VMD) [23] and its multivariate form (MVMD) [24] were proposed in 2014 and 2019 and combined with other artificial intelligence algorithms to achieve load forecasting methods, such as VMD-LSTM [25] and VMD-KMeans-LSTM [26] and so on. However, the above research methods have not fundamentally changed the problem of relying only on the single indicator (historical load), which has great limitations.

Due to the limitations of using a single method, hybrid models and forecasting strategies have gradually received attention in load forecasting due to their high robustness and fast calculation speed. In terms of the combination of optimization algorithms, the bat algorithm can be used in feature selection [27]. The consensus-based mixed-integer particle swarm optimization-assisted TRUST-TECH (CMPSOATT) method can optimize the neural network framework [28]. Moreover, the fruit fly optimization algorithm (FOA), which composes of improved empirical mode decomposition (IEMD), ARIMA and wavelet neural network (WNN) [29], is able to optimize the hybrid model. Finally, feature selection technology and an improved meta-heuristic algorithm are used to construct a two-stage forecasting engine for power load forecasting [30]. In terms of feature expansion, the application of hybrid models are also very common. Convolutional neural network (CNN) is widely used to extract the potential correlations and the contributions to the load [31]. The joint trainings of related cities and communities have achieved more accurate and robust load forecasting results than traditional neural network methods. Unfortunately, most studies are based on the absence of sudden changes (normal days) in the social environment, and whether they can play a role when the sudden changes occur is still blank in the industry.

This paper aims to solve the problems of relying on conventional indicators, huge computational costs, and underutilization of knowledge of related tasks in the above papers, and innovatively consider the impact of the COVID-19 on short-term load forecasting. Specifically, we have made the following contributions:

- To fully use the knowledge of related tasks, this paper implements multi-task learning (MTL) to construct the LSTM-MTL model by adding parameter sharing layers to improve the generalization ability and forecasting accuracy.
- To precisely capture the load changing patterns during the COVID-19, the mobility data (including nine mobility indicators) released by Apple [38] and Google [39] on the COVID-19 pages are innovatively introduced to the model training process to track the rapid changes of load patterns.
- Horizontal and vertical tests in 3 different levels (metropolitan-level, ISO-level, and national-level, including 12 areas) have proved the failure of conventional models in the COVID-19 and the success of the mobility-optimized LSTM-MTL model. It has extreme robustness and can control the forecasting errors of most tasks below 1%. Also, time-based vertical verification shows that LSTM-MTL is a general load forecasting model suitable for different seasons, periods and day types of different regions.
- To solve the black-box problem of load forecasting models, the Shapley additive explanations (SHAP) visual interpretation technology [40] is innovatively incorporated to explain the impact of indicators on load changes in the COVID-19. It visualizes and proves the huge contributions of the mobility indicators on forecasting accuracy and the non-synchronous relationships between the indicator’s correlations and the contributions to the load [41].

2. Preliminaries

2.1. The conventional Short-term load forecasting process for power systems

The purpose of power system short-term load forecasting is to predict the load demand in the sector divided by region or transmission lines for up to 1 week in the future. Fig. 1 shows the brief process of the conventional power system load forecasting process. First, various types of
load-related data (from substations, weather stations, etc.) are imported into the power data centers through different private communication methods. Then, after cleaning and processing the data by the supervisory control and data acquisition (SCADA) system and other facilities, the processed historical load and weather data are sent to the high-performance computing servers to complete the short-term load forecasting process and output the forecasting results. Finally, the power production, transmission, and dispatching departments respond to the power demand of power companies and users based on the load forecasting results at all levels and complete closed-loop managements.

The conventional features used in short-term load forecasting include main features (time-series load and weather, population mobility, etc.) and additional features (economy, policies, etc.). They summarize the changing trends of human production and life from multiple dimensions, thus infinitely approaching the trend of power load forecasting models.

2.2. Long Short-term memory networks

The long short-term memory neural network (LSTM) is a special form of the recurrent neural network (RNN) and performs better in long-term sequence training. The structure of the LSTM is shown in Fig. 2.

The LSTM has introduced the concept of cell state, and three gate structures (i.e., input gate, forget gate, and output gate) are used to maintain and control the flow of timing information. Assume that $C_{t-1}$ and $h_{t-1}$ are the cell state and the hidden layer state at the previous moment. Then, the value of each gate at each time step of the LSTM is shown as (1):

\[
\begin{align*}
 f_t &= \sigma(W_{f} x_t + W_{f} h_{t-1} + b_f) \\
 i_t &= \sigma(W_{i} x_t + W_{i} h_{t-1} + b_i) \\
 c_t &= \tanh(W_{c} x_t + W_{c} h_{t-1} + b_c) \\
 o_t &= \sigma(W_{o} x_t + W_{o} h_{t-1} + b_o) \\
 h_t &= o_t \circ \tanh(c_t)
\end{align*}
\]

where $f_t$ is the forget gate, $i_t$ is the update gate, $c_t$ is the cell candidate state, $c_t$ is the cell state, and $o_t$ is the output gate. The output is calculated by the fully connected layer whose activation function is the sigmoid function and is uniformly represented as a function $\sigma(\cdot)$.

2.3. Load trend analysis in COVID-19

Because of the economic disruption and behavioral restrictions caused by the COVID-19, the electricity consumption in society is entirely different from regular holidays and normal days. Fig. 3 shows the load changes of four locations (New York City, Columbus, Washington, D.C., and Philadelphia) from February 15th to May 15th in 2019 and 2020. Take New York City as an example. On normal days (without COVID-19), there were significant red spikes in the load in New York City, which is consistent with the situation in 2019. However, when the COVID-19 spreads, the daily load changes become smoother than the normal days, and the load spikes disappear. This pattern is the same as in Philadelphia. Besides, in 2020, the overall load level of almost every city has dropped by 10% to 20% compared to 2019, which is contrary to the normal trend (increasing year by year) of urban electric load demand. It is worth noting that since these regions are adjacent in geographic location, we have continuously processed the situation of each independent region (“Region” axis) in this figure. However, readers can also consider the load changes in each region independently according to their preferences.

Many places in the world have issued foot bans, curfews, and travel restrictions to control the spread of the virus. However, the load changes caused by the implementations of these measures are different from those caused by normal holidays. In Fig. 4, we map the changes in power load during the holidays before and during the large-scale outbreak. We use the load data of California on March 15th, 2020 (Sunday, with COVID-19) and March 17th, 2019 (the nearest Sunday one year ago, without COVID-19). During the COVID-19 period in 2020, the power load presents a single peak (evening peak) in the weekend holiday; while in 2019, it presents a regular double peak (morning peak and evening peak) form, which is quite different from the COVID-19 period. Besides, the power load values during the night of the weekend holidays of the COVID-19 period are about 22.7% more than that of the normal weekend holidays, and the power load values during the day are relatively smooth.

Therefore, it is unreliable to consider every day as a holiday during the COVID-19 period. This will cause huge energy waste in the power system and make it difficult for generators to meet the power constraints. Hospitals and public transportation industries that require extremely high reliability of power guarantees are affected by these disadvantages.
3. Mobility optimized LSTM-MTL model

3.1. Mobility

In this paper, mobility data are added to track the trend of human activity during the COVID-19 pandemic. Mobility data can be jointly collected by multiple organizations and enterprises. They can also be obtained by large Internet companies through integration and analysis based on the user’s geographic location information. It quantitatively reflects users’ habits in a region or even a country at a specific moment or time range and is affected by multiple factors such as policies, consumer behavior, and economic activity. Therefore, population mobility is an objective reflection of the mode of social operation.

In this paper, the process of using the mobility data for load forecasting is shown in Fig. 5. The autoregressive hysteresis value of the load is incorporated into the model to track the short-term change trend and stability deviation of the load. The mobility added generalized load forecasting model can be expressed by (2):

$$FL = f(t, d, \text{type}, y_1, y_2, \ldots, y_n, w_1, p_1, m_1, \ldots)$$  \hspace{1cm} (2)

where the dependent features are the following:

- $f(\cdot)$ represents the non-linear mapping between load-related multi-features and forecasting load.
- $FL$ are the forecasting load values for a short time.
- $t \in [0, 1, \ldots, 23]$ are the 24 hours of a day.
- $d \in \{1, 2, \ldots, 365, 366\}$ is the day of a year.
- $\text{type}$ represents the type of day, including legal holidays, weekends, etc.
- $y_1, y_2, \ldots, y_n$ are the main factors, such as the previous power load, fine-grained weather, population mobility and other features directly related to the power load.
- $w_1, p_1, m_1, \ldots$ symbolizes the additional (related) features indirectly related to the power load, such as economic trends, gross domestic product (GDP), etc. These features are selected and used according to different industrial application scenarios.

It is worth noting that we only introduce the generalized methodology of adding population mobility for load forecasting in this section. Readers can directly read Section 4.1 of the experimental part to obtain the details of mobility data.
3.2. Mobility optimized LSTM-MTL model

The power grid is a network that contains multiple regions, and the load provision and consumption in one region will affect the load provision and consumption in other regions. The population flow between regions also follows the same pattern. Therefore, it is of great significance to use the correlation of load information between multiple regions at the same level to expand the data scale.

In this paper, multi-task learning (MTL) achieves the goals shown in Fig. 6. The information used in the load forecasting task of each region comes not only from the original data information inputs but also from the information of other regional tasks obtained through knowledge transfer. The process can be expressed by (3):

\[ FL_i = M(t, d, FO_i, FT_i) \quad \forall i \in (1, 2, ..., n) \]  

where \( FL_i, FO_i, FT_i \) represent the \( i \)-th forecasting task, the own knowledge of the \( i \)-th task, and the knowledge transferred from other related tasks, respectively. \( n \) is the number of forecasting tasks that need to be trained.

Two forms of the training process are considered. One process is single-task learning, the dataset (with or without mobility) of the target task area directly passes through several hidden layers and finally outputs the load forecasting results of the area. The other process is multi-task learning. The LSTM-MTL model based on mobility optimization is proposed and shown in Fig. 7. Firstly, the tasks (Task 1, Task 2, ..., Target task, ..., Task n) are divided into related tasks and target tasks (can be flexibly adjusted). The load datasets (with or without mobility data) of the area adjacent to the related tasks are mainly used. The target tasks use the load datasets of the target areas. Then, the load-related data of multiple locations are firstly compressed through the normalization layer and the dense layers for normalization and data dimensionality reduction. The LSTM parameters sharing layer combines all the tasks (related tasks and target tasks) to achieve parallel knowledge transfer and noise balance between tasks, thereby using the hard sharing mechanism to implement multi-task learning. Then, each task enters its fully connected layers and finally reaches output layers, outputting the forecasting results of the related area and the target area. Many studies focusing on training time have proved that multi-task learning can greatly reduce training time and provide the possibility to implement large-scale, multi-region, and rapid load forecasting. It is worth noting that, although we set the target task and related tasks in this paper, the sub-tasks of multi-task learning are always parallel. That is, there is no definite primary and secondary task like transfer learning. Readers can arbitrarily set the task of the region they are concerned with as the target task.

Besides, for the completeness of knowledge and theory, we need to emphasize that since the optimization of the internal parameters of the neural network is a non-deterministic polynomial complete (NPC) problem. It is impossible to find the completely optimal parameters within tolerable time to adapt to all load forecasting scenarios. Therefore, we can only obtain a neural network that can adapt to most (not all) load forecasting tasks during COVID-19 as much as possible through heuristic neural network training, and its internal parameters are always sub-optimal.

3.3. Shapley additive explanations

Load forecasting models are typical black-box models. For a neural network model constructed from a dataset of multiple load-related features, how each feature sensitively affects the output of load task area directly passes through several hidden layers and finally outputs the load forecasting results of the area. The other process is multi-task learning. The LSTM-MTL model based on mobility optimization is proposed and shown in Fig. 7. Firstly, the tasks (Task 1, Task 2, ..., Target task, ..., Task n) are divided into related tasks and target tasks (can be flexibly adjusted). The load datasets (with or without mobility data) of the area adjacent to the related tasks are mainly used. The target tasks use the load datasets of the target areas. Then, the load-related data of multiple locations are firstly compressed through the normalization layer and the dense layers for normalization and data dimensionality reduction. The LSTM parameters sharing layer combines all the tasks (related tasks and target tasks) to achieve parallel knowledge transfer and noise balance between tasks, thereby using the hard sharing mechanism to implement multi-task learning. Then, each task enters its fully connected layers and finally reaches output layers, outputting the forecasting results of the related area and the target area. Many studies focusing on training time have proved that multi-task learning can greatly reduce training time and provide the possibility to implement large-scale, multi-region, and rapid load forecasting. It is worth noting that, although we set the target task and related tasks in this paper, the sub-tasks of multi-task learning are always parallel. That is, there is no definite primary and secondary task like transfer learning. Readers can arbitrarily set the task of the region they are concerned with as the target task.

Besides, for the completeness of knowledge and theory, we need to emphasize that since the optimization of the internal parameters of the neural network is a non-deterministic polynomial complete (NPC) problem. It is impossible to find the completely optimal parameters within tolerable time to adapt to all load forecasting scenarios. Therefore, we can only obtain a neural network that can adapt to most (not all) load forecasting tasks during COVID-19 as much as possible through heuristic neural network training, and its internal parameters are always sub-optimal.

3.3. Shapley additive explanations

Load forecasting models are typical black-box models. For a neural network model constructed from a dataset of multiple load-related features, how each feature sensitively affects the output of load

---

**Fig. 6.** The graph of knowledge transfer in multi-task learning. The graph only shows the process of knowledge transfer between two regions.

---

**Fig. 7.** The training process of mobility optimized LSTM-MTL model. Mobility data is adopted to the training process of the LSTM-MTL model.
forecasting is still a problem to be solved. Reasonable model interpretation is extremely significant for improving the effect of model retraining and analyzing abnormal samples. Besides, in some scenarios, transparency and interpretability are essential criteria for trustworthy machine learning models, such as finance and risk control. Model explanations based on strict mathematical foundations are needed to verify validity for the newly proposed models and features.

In this paper, the Shapley additive explanations (SHAP) is introduced to explain the output of the proposed LSTM-ML model and show the real effect of load-related mobility features. The whole process is shown in Fig. 8. The SHAP technology connects the optimal credit allocation with local explanations using the classical Shapley values from game theory and their related extensions [40]. The explanation technology regards all input load-related feature quantities as “contributors” of the results, and the degree of contribution will affect the increase or decrease of the results. The expression is as (4):

\[ y_t = y_{base} + f(x_{i1}) + f(x_{i2}) + \ldots + f(x_{i6}) \]  

where \( y_{base} \) represents the mean value of the forecasting load values, \( f(x_{i1}) \) represents the Shapley value (contribution value) of the first load-related feature in the \( i \)-th sample to the load forecasting result \( y_t \). If \( f(x_{i1}) \) is greater than 0, it means that the first feature positively affects the predicted load result \( y_t \) and increases \( y_t \) from \( y_{base} \); otherwise, it has a negative effect and makes it decrease.

To facilitate the SHAP conveniently, we use a third-party tool on GitHub [44]. After our model training is completed, we perform visual interpretation and analysis of our proposed model to better understand each load-related feature’s influence direction and degree of influence on the load forecasting value at a specific point of time. Besides, it can be additionally noted that, according to the descriptions of papers [45] and the website of this third-party tool. The SHAP tool can be applied to our forecasting model with LSTM as the forecasting engine and is widely used in other machine learning models, such as CNN and tree ensemble methods (XGBoost, LightGBM, etc.).

4. Cases for short-term load forecasting in COVID-19

4.1. Data collection and simulation platform

This test can be performed online (Google Colab, etc.) or offline (local computers or servers). To facilitate training conveniently, the offline method is used in this study. Models are trained on a laptop with the Intel-9750H® processor, RTX2060 graphics card, TensorFlow 2.4 deep learning framework, Keras 2.4, and Windows 10 64-bit system.

The data sources of this study are as follows:

1) Load data

Table 1

| Data level          | Source                        |
|---------------------|-------------------------------|
| Nation-level        | ENTSO-E                       |
| ISO-level           | USA EIA                       |
| Metropolitan-level  | PJM, NYISO and other public institutions |

Table 2

| From       | Number of features | Feature contents                      |
|------------|--------------------|---------------------------------------|
| Apple      | 3                  | Driving, Walking, Transit             |
| Google     | 6                  | Retail and recreation, Grocery and pharmacy, Parks, Transit stations, Workplaces, Residential |

Power practitioners of different levels need different levels of power load data. The nation-level data (Switzerland, France, Germany, Italy), ISO-level data (APS, CAISO, NEVP, PACE) and the metropolitan-level data in the eastern United States (New York, Columbus, Philadelphia, Washington D.C.) are used in this study. The national-level data are from ENTSO-E [46], the ISO-level data can be obtained from U.S. EIA [47], and the metropolitan-level data are mainly obtained from the PJM [48], NYISO [49], and several public institutions, as shown in Table 1.

2) Mobility data

The mobility data of various countries and regions are obtained through the COVID-19 pages of Apple and Google and have nine dimensions. In this paper, the mobility data of the United States during the COVID-19 pandemic is obtained. The specific mobility features are shown in Table 2.

3) Fine-grained time data

Because the load forecasting model is a typical time-series model, the fine-grained time features need to be added for training an accurate model. In this paper, time is divided into 32 dimensions for representation, of which 24 dimensions are used to represent 24 hours (\( H_1, H_2, \ldots, H_{24} \)) of a day, seven dimensions are used to represent seven days (\( Sun, Mon, \ldots, Sat \)) of a week, and the remaining one dimension is used to represent holidays.

4) Weather data

The hourly weather data, including multiple weather features, is needed and collected from the WWO’s online API [50]. Besides, the weather data of the ISO-level and national-level datasets are mainly taken from the central city of the ISOs and the capital city of the countries. To better characterize the influence of regional weather on the load, five dimensions of weather features, including cloud-cover, humidity, rainfall, air pressure, and temperature, are mainly chosen because many scholars have proved that they often have a potentially huge impact on load in their previous studies [42].

4.2. Model settings

To better compare the effects of our models and conventional models’ responses during the COVID-19 period, some model settings to facilitate comparison are shown in Table 3. Task types are divided into single-task learning and multi-task learning. For feature types, T, W, L, and M represent time, weather, historical load, and mobility, respectively. Besides, to compare the impact of data in different time ranges on the forecasting results, the data from January 1st, 2018 to December 31st, 2019 is defined as the “old” data, and the data from February 15th, 2020 to August 31st, 2020 is defined as the “new” data.
2020 to August 15th, 2020 is defined as the “new” data. It is worth noting that, in addition to the above model, the accuracy of the forecasting results of other non-sequential mainstream machine learning algorithms in “new” datasets (time, weather, load, population mobility) are tested, such as extreme gradient boosting (XGBoost), random forest regression (RFR), support vector regression (SVR), multi-layer perceptron regression (MLR), and light gradient boosting machine (LGBM). The datasets used are precisely the same as the previous seven models.

The datasets are divided into the training sets, the validation sets, and the test sets. Considering that the Adam optimizer has better performance than the SGD optimizer, the Adam optimizer is applied to all model training processes. The basic settings of some parameters of the datasets and the model training processes are shown in Table 4. As for forecasting horizons, there are different forecasting horizons for the “new” dataset and the “old” dataset, but both are the last week of the corresponding dataset. Specifically, for the “new” dataset, the forecasting horizon is from 0 a.m. on August 9th, 2020 to 11 p.m. on August 15th, 2020. There are a total of 168 load values (168 steps) that need to be predicted; for the “old” dataset, the forecasting horizon is from 0 a.m. on December 25th, 2019 to 11 p.m. on December 31st, 2019, which is also 168 steps.

Besides, there are many methods for evaluating accuracy. Standard methods are mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). In this paper, the MAPE is used to evaluate the accuracy of the load forecasting model. The calculation method of the MAPE value \( L_{MAPE} \) is:

\[
L_{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{L_t - \hat{L}_t}{L_t} \right| \quad L_{MAPE} \in [0, 1]
\]  

(5)

Among them, the value \( L_{MAPE} \) represents the percentage of the error between the forecasting load and the actual load, \( N \) represents the size of the batch size, \( \hat{L}_t \) represents the forecasting load value at the time \( t \), the unit is \( MW \), \( L_t \) represents the actual load value at the time \( t \).

Besides, to overcome the one-sided defect when a single evaluation indicator is used to evaluate the results of load forecasting, we also introduce the Theil inequality coefficient (TIC) in statistics and analyze the load forecasting results from the perspective of the variance of the forecasting value and the actual value. The calculation method of the TIC is shown in (6):

\[
TIC = \frac{1}{N} \sum_{i=1}^{N} \left( \hat{L}_i - L_i \right) \quad TIC \in [0, 1]
\]

(6)

where TIC represents the calculation result of the Theil inequality coefficient. The definition of other symbols is the same as (5). In our actual simulation, to display the results more intuitively, we multiply the obtained MAPE and TIC values by 100. The MAPE value multiplied by 100 has the meaning of conversion percentage, but TIC does not.

4.3. The padding of missed data and the correlation check of mobility data

4.3.1. The padding of missed data

In this test, due to the high degree of completeness of the load data, there is no need to supplement the missing values for the power load data. However, according to the note published by Apple on the COVID-19 page, the data of the three major population mobility indicators (driving, transit, and walking) on May 11th and May 12th, 2020 were not recorded in time due to system maintenance. Therefore, measures need to be taken to supplement the missing values of the data for these two days.

Specifically, we trained a small LSTM neural network, using Google’s data on May 11th and May 12th, 2020 as input to predict the three major mobility values that Apple has missed in these two days. The process of filling in missing values can be expressed as (7):

\[
M = \text{padding}(G_1, G_2, ..., G_d) \quad x \in [1, \text{len}(G)]
\]

(7)

where \( M \) represents the prediction of missing values, padding represents the neural network used to fill the missing values, we will express it as a functional form of padding(.), \( G_1, G_2, \text{etc.} \) represent the data input used to train the neural network. In this test, the mobility data from Google's (6 indicators) are the input data. Besides, it is worth noting that although we only used existing mobility data obtained from other channels to fill in the missing population mobility data, other types of missing data (such as some partial missing load data) can also be filled by similar ways.

4.3.2. The correlation check of mobility data

Training the power load forecasting model with highly correlated indicators with the power load will improve the forecasting accuracy and vice versa. Therefore, it is necessary to check the correlation of mobility data before training.

The temperature indicator has been proven effective in almost all previous related studies and has a profound impact on power load [51]. Therefore, we use the correlation coefficient between temperature indicators and power load as a reference to see whether the mobility indicators have the same or even higher degree of relevance to illustrate the potential values of mobility indicators. In our tests, the Spearman’s rank correlation coefficient [52] is employed as a tool to verify the correlation between load and mobility. The calculation formula of the Spearman’s rank correlation coefficient is in (8):

\[
S = \frac{\sum_{i=1}^{n}(L_i - \overline{L})(M_i - \overline{M})}{\sqrt{\sum_{i=1}^{n}(L_i - \overline{L})^2 \sum_{i=1}^{n}(M_i - \overline{M})^2}} = 1 - 6 \sum_{i=1}^{n} \frac{d_i^2}{n(n^2 - 1)}
\]

(8)

where \( S \) represents the calculation result of the Spearman’s rank correlation coefficient statistic. \( L \) represents the actual electrical load value during the entire 6-month period since the outbreak of the COVID-19, \( \overline{L} \) and \( \overline{M} \) represent the average value of the electrical load and a single mobility indicator. \( d_i \) is the difference of rank, \( n \) is the number of the
data.
The \( S \) value of power load and population mobility characteristics and the corresponding group curve diagrams for the 6 months from February 15\(^{th} \) to August 15\(^{th} \) of 2020 are shown in Table 5 and Fig. 9.

Taking the \( S \) value of the temperature indicator and the power load as a reference, 5 of the 9 human mobility indicators have a stronger correlation than the temperature indicator, and the absolute values of the remaining 4 indicators are also close to the \( S \) value of the temperature indicator. This shows that during the COVID-19 period, mobility indicators have a close or even greater potential ability to follow changes in power load than temperature indicators. Among them, the \( S \) values of driving indicator, transit indicator are as high as 0.846 and 0.724. Besides, except for the negative \( S \) value of the residential indicator, the \( S \) values of the other indicators are all positive. It shows that the behavior of living at home will tend to reduce the electrical load, which is consistent with the actual situation during the COVID-19. Therefore, incorporating population mobility indicators into the training of power system load forecasting has great potential value for improving the load forecasting accuracy of the power system.

### 4.4. Evaluation results

#### 4.4.1. The validity verifications of the models and indicators in COVID-19

To better train the load forecasting model in Section 4.2, the load forecasting task of New York City is selected as the target task, and tasks of the Washington D.C., Columbus, Philadelphia are selected as the related tasks to support the load forecasting task of New York City. Besides, when testing the time of parallel tasks, the data of the Boston area is included.

One of the biggest differences between single-task learning and multi-task learning is the length of calculation time. Table 6 shows the training time of each region using new mobility optimized data. It can be seen that the calculation time of each region is very close, ranging from 96 s to 100 s. There is no significant difference in training time between locations for single-task learning. Table 7 shows the impact of different numbers of parallel load forecasting tasks on training time and speed-up. When the number of parallel tasks is small, the speed-up effect of multi-task learning is not obvious, with a speed-up rate of 14.84%. Surprisingly, although the parallel training of load forecasting models in more locations will increase the training time, the speed-up rate obtained relative to single-task learning is monotonically increasing. When the

---

**Table 5**

| Mobility indicators | Temperature | Driving | Transit | Walking | Retail and recreation | Grocery and pharmacy | Parks | Transit stations | Residential | Workplaces |
|---------------------|-------------|---------|---------|---------|-----------------------|----------------------|-------|-----------------|------------|------------|
| \( S \)              | 0.620       | 0.846   | 0.724   | 0.691   | 0.684                 | 0.530                | 0.594 | 0.638           | -0.539     | 0.461      |

Note: in this table, the temperature indicator serves as a reference.

---

![Fig. 9](image-url)  
*Fig. 9. The Spearman correlation verification group graphs of New York City’s population mobility indicators and electricity load during the COVID-19. (A) Electricity load and temperature. (B) Electricity load and driving. (C) Electricity load and traffic transit. (D) Electricity load and walking. (E) Electricity load and retail and recreation. (F) Electricity load and grocery and pharmacies. (G) Electricity load and parks. (H) Electricity load and transit stations. (I) Electricity load and residential. (J) Electricity load and workplaces. Almost all population mobility indicators have reached a level that is more than moderately related to power load and can follow sudden changes of power load in time in the COVID-19.*
number of parallels is 3, the speed-up rate reaches 50%. Furthermore, when the number of parallel training locations for multi-task learning is 5, the speed-up rate is close to 100%. This shows that the multi-task load forecasting method proposed in this study can significantly reduce the training time of a load forecasting model in multiple locations. If the training scale is larger, the acceleration effect will be more significant. Therefore, the method proposed in this study is suitable for large-scale load forecasting systems.

The load forecasting results of different models in New York City from August 9th to 15th are shown in Fig. 10, Fig. 11, Table 8 and Table 9. Whether using the LSTM-Old model of LSTM or the NN-Old model of traditional RNN, as long as the load forecasting model is based on historical data over the years, the MAPE value will exceed 5%. The deviation between the predicted value and the actual value of some predicted positions is as high as 1000 MW. Therefore, models based on large historical datasets are not reliable during the COVID-19. Besides, although the error of the LSTM-Retrain-New model trained with a "new" dataset is not as large as the "old" dataset used to train the NN-Old model, the addition of mobility data can make the both model successfully follow the load patterns and control the errors below the tolerable values in the COVID-19. The LSTM-MLT model we proposed has the lowest forecasting error (only 0.467%, which is 92.3% lower than NN-Old), which is far better than other traditional models (such as GRU, XGBoost, etc.). Besides, it shows that the use of multi-task learning to realize cross-task knowledge transfer is helpful to load forecasting. Fig. 12 shows the error distributions of the four LSTM-based model. The error deviation ratio value of LSTM-MLT clusters around zero with a high probability, which is much better than other LSTM-based models' results. This continues to confirm the high precision of the LSTM-MLT model during the COVID-19 period. When we conduct a horizontal comparison in Table 8 and Table 9 of the training effects and forecasting effects of the load forecasting models trained on datasets of different scales, we can find that our proposed LSTM-MLT model have all achieved the highest forecasting accuracy in 12 forecasting tasks on three scales during the COVID-19. However, the accuracy of forecasting tasks of different scales is entirely different. The forecasting MAPE values of the four cities in the United States are between 0.467% and 0.756%, and the 100 times TIC value is concentrated between 0.311 and 0.496; the forecasting MAPE values of the four ISO levels are concentrated between 0.453% and 0.807%, and the 100 times the TIC value is concentrated between 0.287 and 0.534; the forecasting MAPE values of the four European countries are focusing between 0.388% and 0.688%, and the TIC values are concentrated between 0.233 and 0.617. Therefore, the forecasting accuracy and numerical stability of metropolitan-level load forecasting are higher than that of ISO-level load forecasting. This may be caused by other factors that cause system errors and data noise, and the mobility data and weather data considered in the simulation do not cover all possible fine-grained areas. Taking the results of the proposed LSTM-MLT model as an example, the MAPE values of metropolitan-level load forecasting tasks are generally lower than 0.7%, and the MAPE values of ISO-level load forecasting tasks are usually higher than 0.8%. However, they are still good forecasting results. When the nation-level load forecasting tasks are carried out, we are surprised to find that the accuracy and numerical stability of

![Fig. 10. The load forecasting curves of the traditional RNN model using the dataset of New York City. Only the NN-Mob-Single model that uses mobility data has better ability to follow the actual load.](image1)

![Fig. 11. The load forecasting curves of the LSTM and LSTM-MLT model using the dataset of New York City. The LSTM-MLT model almost perfectly follows the actual load changes, while the LSTM-Retrain-New model, LSTM-Mob-Single model and LSTM-Old model have poor following ability.](image2)
Table 8
The Validation (MAPE And TIC) Results Of Load Forecasting In Model Training.

| MAPE(%) (TICx100) | NY | COL | WDC | PHIL | CAISO | APS | NEVP | PACE | SWIT | FRA | ITA | GER |
|--------------------|----|-----|-----|------|-------|-----|------|------|------|-----|-----|-----|
| **NN-Old**         | 6.855 | 5.177 | 6.284 | 6.215 | 8.421 | 7.488 | 5.568 | 9.017 | 5.338 | 4.992 | 5.337 | 5.437 |
| (3.803)            | (3.120) | (3.678) | (3.605) | (5.427) | (4.671) | (2.994) | (6.567) | (4.076) | (3.734) | (4.073) | (4.310) |
| **NN-Retrain-New** | 5.740 | 8.551 | 10.496 | 9.535 | 10.063 | 9.603 | 9.089 | 10.578 | 4.236 | 4.515 | 4.840 | 3.785 |
| (3.469)            | (5.197) | (6.113) | (5.700) | (6.757) | (5.560) | (4.686) | (7.418) | (2.887) | (3.097) | (3.063) | (2.162) |
| **NN-Mob-Single**  | 1.001 | 1.349 | 1.423 | 1.312 | 2.191 | 3.321 | 3.210 | 2.099 | 1.574 | 1.408 | 1.829 | 1.017 |
| (0.740)            | (1.127) | (1.452) | (1.025) | (1.931) | (2.770) | (2.146) | (2.101) | (1.114) | (0.946) | (1.356) | (0.688) |
| **LSTM-Old**       | 1.409 | 1.446 | 1.505 | 1.466 | 1.397 | 2.386 | 2.018 | 1.445 | 2.348 | 2.147 | 2.347 | 1.994 |
| (0.985)            | (0.994) | (1.007) | (1.080) | (0.951) | (1.563) | (1.350) | (0.978) | (2.115) | (1.444) | (2.115) | (1.768) |
| **LSTM-Retrain-New** | 1.448 | 1.273 | 2.373 | 2.083 | 7.562 | 7.235 | 3.716 | 4.993 | 1.532 | 1.612 | 1.555 | 1.702 |
| (0.910)            | (8.741) | (1.647) | (1.370) | (4.762) | (4.796) | (2.367) | (3.173) | (1.001) | (1.109) | (1.054) | (1.157) |
| **LSTM-Mob-Single** | 1.723 | 1.148 | 1.365 | 1.283 | 1.320 | 3.307 | 2.468 | 1.607 | 1.269 | 1.555 | 1.356 | 1.892 |
| (1.016)            | (0.926) | (1.373) | (0.976) | (1.040) | (2.288) | (1.796) | (1.621) | (0.856) | (1.073) | (0.997) | (1.106) |

Table 9
The Accuracy (MAPE And TIC) Results Of Load Forecasting In Model Testing.

| MAPE(%) (TICx100) | NY | COL | WDC | PHIL | CAISO | APS | NEVP | PACE | SWIT | FRA | ITA | GER |
|--------------------|----|-----|-----|------|-------|-----|------|------|------|-----|-----|-----|
| **NN-Old**         | 6.906 | 5.236 | 6.503 | 6.316 | 8.652 | 7.849 | 7.861 | 8.052 | 5.495 | 5.213 | 5.548 | 5.568 |
| (3.942)            | (3.256) | (4.023) | (3.845) | (5.968) | (4.967) | (5.213) | (5.561) | (5.987) | (5.023) | (5.463) | (5.688) |
| **NN-Retrain-New** | 6.023 | 7.521 | 9.213 | 9.832 | 10.865 | 9.578 | 11.012 | 10.126 | 4.996 | 4.679 | 5.254 | 4.028 |
| (3.968)            | (5.231) | (5.984) | (5.843) | (7.237) | (7.964) | (7.568) | (7.021) | (3.568) | (3.223) | (3.698) | (4.328) |
| **NN-Mob-Single**  | 1.298 | 1.652 | 2.099 | 1.963 | 2.737 | 2.928 | 4.144 | 3.007 | 1.646 | 1.482 | 1.759 | 1.203 |
| (0.854)            | (1.256) | (1.442) | (1.301) | (1.718) | (1.856) | (2.699) | (2.045) | (1.115) | (1.901) | (1.325) | (0.808) |
| **LSTM-Old**       | 6.176 | 5.701 | 7.319 | 7.365 | 8.150 | 8.084 | 7.976 | 9.401 | 1.563 | 1.766 | 1.563 | 5.429 |
| (5.552)            | (4.379) | (4.323) | (3.975) | (5.580) | (5.118) | (2.367) | (6.665) | (1.318) | (1.254) | (1.318) | (4.333) |
| **LSTM-Retrain-New** | 2.131 | 2.511 | 3.154 | 3.060 | 3.350 | 5.191 | 4.148 | 3.557 | 1.834 | 1.912 | 2.056 | 1.604 |
| (1.406)            | (1.726) | (2.116) | (2.045) | (3.172) | (4.131) | (2.892) | (3.012) | (1.125) | (1.223) | (1.667) | (0.989) |
| **LSTM-Mob-Single** | 1.744 | 1.608 | 2.048 | 1.602 | 1.271 | 2.162 | 2.107 | 1.609 | 1.689 | 1.771 | 1.453 | 2.143 |
| (1.071)            | (1.105) | (1.382) | (1.019) | (1.146) | (1.397) | (1.746) | (1.101) | (1.254) | (1.099) | (1.931) | (1.522) |
| **LSTM-MTL**       | 0.647 | 0.579 | 0.756 | 0.622 | 0.801 | 0.807 | 0.788 | 0.453 | 0.688 | 0.884 | 0.787 | 0.388 |
| (0.311)            | (0.376) | (0.496) | (0.396) | (0.521) | (0.496) | (0.534) | (0.287) | (0.457) | (0.617) | (0.557) | (0.233) |

Note: the name of cities and countries are abbreviated. Specifically, NY = New York, COL = Columbus, WDC = Washington D.C., PHIL = Philadelphia, SWIT = Switzerland, FRA = France, ITA = Italy, GER = Germany.
the nation-level tasks are higher than that of the ISO-level tasks and lower than that of the metropolitan-level tasks. This is probably because the country’s overall load value is more resistant to disturbances and closer to a stable time series. In general, the LSTM-MTL model we proposed can achieve good results when dealing with different levels of load forecasting tasks.

As for other conventional models, such as XGBoost, RFR, etc., we used particle swarm optimization (PSO) during trainings to optimize their parameters to maximize their ability to adapt to datasets of various regions. Specifically, most MAPE values are greater than 5% (generally acceptable load forecasting results are less than 3%), and the ISO-level prediction accuracy and numerical stability are the worst. The MAPE values of SVR and MLR when forecasting each ISOs' load are as high as 11% to 13%, which is an awful result. This also proves the effectiveness of our proposed LSTM-MTL model. Readers can refer to the Section B of the Appendix for more detailed information of model parameters.

4.4.2. Time and season validity verification

In this section, we will consider the impact of seasonal and time factors superimposed on the changes brought about by the COVID-19. We continue to obtain newer data for each dataset from February 15th, 2020 to August 15th, 2021 from the data sources introduced in Section 4.1 to carry out further longitudinal comparison experiments about different periods and seasons to fully demonstrate the effectiveness and capabilities of the LSTM-MTL model. It is worth noting that the time from February 15th, 2020 to August 15th, 2021 covers all stages and phases of the COVID-19, such as the initial emergence of the COVID-19, the outbreak, the suppression, and becoming the focus of normalization of prevention and control (towards normal life). Therefore, the verification conditions will be more stringent.

In these tests, we keep the model settings, the time span of each sub-dataset, and other factors unchanged and use a sliding time window to verify the results of the load forecasting tasks in the middle of each month. For example, to predict the load from May 9th to May 15th of 2021, we use the data from November 15th, 2020 to May 1st, 2021 as the training set; the data from May 2nd to May 8th of 2021 as the validation set. In general, the time span of sub-datasets is always half a year.

The sliding forecasting verification results output by the LSTM-MTL models for 12 related locations, ISOs and countries from August 15th, 2020 to August 15th, 2021 (under the influence of the COVID-19) are shown in Table 10. Most regions, ISOs, and countries can maintain a good forecasting error (MAPE values lower than 1%) using the LSTM-MTL models in most months. The highest error comes from CAISO in the load forecasting tasks of December of 2020, and the MAPE value reached 1.761%. It can also be observed that the forecasting errors of some cities and ISOs during the summer of 2020 (August) and 2021
In general, the LSTM-MTL model we proposed can follow the changes in power load well in both situations (the impact of COVID-19 is more significant, and social production and life tend to be normal) and has good versatility and adaptability. However, electric power workers should pay special attention to the load dispatch control during the changing seasons.

4.5. Model interpretation

Fig. 13 to Fig. 15 show the use of SHAP visual interpretation technology to explain the contribution of each feature of the LSTM-MTL model trained with the mobility data of New York in Section 4.4.1. In Fig. 13, in the absence of COVID-19 (the normal day before the ban), the electricity load value in New York City is relatively high, almost all travel indicators tend to increase (red line) the electrical load.

Fig. 13. The SHAP graph of different features in New York City on February 21st, 2020 (8 a.m., without COVID-19). The value of each indicator in the graph is the value in the dataset at this moment. For details, please refer to the description of the dataset of this test in Section A of Appendix. The more red or blue area (the longer the line) an indicator occupies indicates the greater the absolute Shapley value of this indicator, that is, the stronger the impact on the load.

Fig. 14. The SHAP graph of different features in New York City on April 24th, 2020 (8 a.m., with COVID-19). The specific meaning of the elements in the graph is the same as Fig. 13.
Fig. 15. The distribution graph of Shapley values of different features in the dataset of New York City. Only the top 20 most influential indicators are taken to present in the graph.

5. Conclusion

In this paper, the LSTM-MTL model based on mobility optimization is proposed to perform more accurate short-term load forecasting during the COVID-19 pandemic. The conclusion of this paper is as follows:

• The multi-task learning method allows knowledge transfer in parallel tasks to improve the generalization ability of each task and the quality of the load forecasting model, which significantly reduces the model training time and improves the forecasting accuracy.
• The adoption of mobility data dramatically improves the real-time ability of the load forecasting model to capture drastic changes in load patterns during the COVID-19 period.
• By using the SHAP model interpretation technology, and through horizontal verification and vertical verification based on different models and different time periods, the LSTM-MTL model and mobility indicators have been proved effective in helping to greatly improve the accuracy of load forecasting tasks during the COVID-19.

Future work includes the use of multiple neural network models (such as convolutional neural network), multiple weight distribution methods (such as attention mechanism) of load data to build a large-scale and comprehensive forecasting model. Besides, the potential of using this paper’s methods and population indicators to predict load in other types of social changes needs to be further explored, such as the Great Strike, the Great Depression, etc.

CRediT authorship contribution statement

Jiefeng Liu: Writing – original draft, Writing – review & editing. Zhenhao Zhang: Writing – original draft, Writing – review & editing. Xianhao Fan: Investigation, Supervision. Yiyi Zhang: Visualization, Project administration, Funding acquisition. Jiaqi Wang: Visualization, Project administration, Funding acquisition. Ke Zhou: Resources, Software. Shuo Liang: Resources, Software. Xiaoyong Yu: Conceptualization, Software, Data curation. Wei Zhang: Conceptualization, Software, Data curation.

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.
Acknowledgements

This material was supported by the Guangxi Bagui Young Scholars Special Funding, Guangxi Key Laboratory of Power System Optimization and Energy Technology Project, Award Number AE3020001829.

Appendix

A Dataset example

To facilitate readers to grasp the structure of the datasets used in the paper, we intercepted part of the data as an example of dataset construction to enable readers to build their dataset and quickly migrate the methods proposed in this paper to their project. Part of the data of the dataset in this paper is shown in Table A.1.

The indicators used in our datasets are power load indicators, weather indicators, time indicators, day type indicators, and population mobility indicators. The power load indicator is a specific value collected from various agencies, and the unit is unified as megawatts (MW). There are 5 weather indicators, and their units follow the International System of Units. We all use one-hot vectors to express the value of time indicator. For example, the first time point in Table A.1 is 1:00 AM on Saturday, February 15th, 2020. Then we set “Hour 1 = 1”, and another hour indicator such as “Hour 2” and “Hour 3” are 0; set “Saturday = 1”, and the indicator of the week such as “Sunday” and “Monday” is 0. Since we are implementing short-term load forecasting in this paper, it does not make sense to incorporate monthly indicators into training, so we do not consider including monthly time indicators. The values of Apple’s population mobility indicator are determined based on different criteria from Google. Apple’s population mobility indicator values (driving, walking, transit) on a certain day are based on January 13th, 2020 (baseline value is 100); Google’s baseline values are the median values of the 5-week period from January 3rd to February 6th, 2020. On other days, calculate the floating percentage values based on these baseline values as the population mobility values.

Due to space reasons, we use multiple rows in Table A.1 to show the data changes at each moment. But, to facilitate model training, all of our data for each data point (time point) is displayed in one row, that is, 1 row and 47 columns.

B Parameters tuning

To maximize the predictive power of these non-time series and ordinary models, it is necessary to tune their parameters, which in turn proves the performance of our model. We used the PSO algorithm to optimize the parameters of these models. The details of the PSO algorithm can be seen in [53].

To make these models universal, our adjusted parameters may not perform the best in a certain region or a country, but the average MAPE value is the smallest. The parameters of the non-time series model and the normal model after PSO optimization are shown in Table B.1. It is worth noting that optimizing each parameter using an optimization

### Table A.1

The Structure of Dataset.

| Time       | Load  | Cloud-cover | Humidity | Precipitation | Pressure | Temperature | Sunday | Monday | Tuesday | Wednesday |
|------------|-------|-------------|----------|---------------|----------|-------------|--------|--------|---------|-----------|
| 2020.02.15| 5269.1| 0           | 43       | 0             | 1035     | –4          | 0      | 0      | 0       | 0         |
| 01:00:00   |       |             |          |               |          |             |        |        |         |           |
| Saturday   |       |             |          |               |          |             |        |        |         |           |
| Thursday   | 0     | 0           | 1        | 0             | 1028     | –4          | 0      | 0      | 0       | 0         |
| 08:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| Hour 8     | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 09:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 10:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 11:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| Walking    | 144.7 | 3           | –4       | –2            | –4       | 1           | 0      | 0      | 0       | 0         |
| 2020.02.16 | 4929.5| 59          | 64       | 1             | 1028     | 0           | 0      | 0      | 0       | 0         |
| 02:00:00   |       |             |          |               |          |             |        |        |         |           |
| Sunday     | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 08:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| Hour 8     | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 09:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 10:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 11:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| Walking    | 124.04| 7           | 3        | 13            | 1013     | –1          | 0      | 0      | 0       | 0         |
| 2020.07.04 | 6153.818| 44       | 69       | 0             | 1013     | 0           | 0      | 0      | 0       | 0         |
| 11:00:00   |       |             |          |               |          |             |        |        |         |           |
| Saturday   | 0     | 0           | 1        | 0             | 1013     | 0           | 0      | 0      | 0       | 0         |
| Independence Day | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 08:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| Hour 8     | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 09:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 10:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| 11:00:00   | 0     | 0           | 0        | 0             | 0        | 0           | 0      | 0      | 0       | 0         |
| Walking    | 70.16 | –44         | –6       | 129           | –43      | –31         | 0      | 0      | 0       | 0         |
algorithm will be a process of high computational cost, and it often takes a long time on devices with fewer computing resources. Therefore, we encourage readers to focus more on data collection and processing, rather than relying too much on model parameters, that is, follow the data-centric rather than model-centric philosophy.

C SHAP interaction explanation

Due to the unique characteristics of the Shapley value, SHAP visual interpretation technology can not only explain the interaction between load and a specific feature, but can also interpret the relationship between load and two features. That is, three related values (power load and two features) and their corresponding influence directions can be seen in one graph at the same time. As an extended content, this content can make it easier for readers to have a deeper understanding of the application of SHAP visual interpretation technology and can also be used as an application example for readers to apply SHAP to other fields of their own.

We continue to expand the discussion of the model interpretation process in Section 4.5. Fig. C.1 shows the relationship between the temperature indicators and the three most temperature-related indicators (transit indicator, driving indicator, and retail and recreation indicator) in the load forecasting model of New York City during the COVID-19 period. The value of temperature has varying degrees of influence on people’s travel patterns and the retail industry’s development, which indirectly affects the changes in load. It can be seen from figure (a) and figure (c) that the lower temperature promotes people’s travel more, and is also more conducive to the development of retail and

| Table B.1 | The Key Parameter Details of Non-time Series and Normal Model. |
|-----------|-------------------------------------------------------------|
| Model name | Parameter details                                           |
| GRU       | epochs (PSO) 150, optimizer adam, loss mean_squared_error, activation relu |
| XGBoost   | n_estimators (PSO) 550, max_depth (PSO) 5, criterion mse, max_depth (PSO) 3 |
| RFR       | n_estimators (PSO) 150, n_jobs 1, criterion mse, max_depth (PSO) 3 |
| SVR       | C 850, gamma (PSO) 1, kernel mse, max_iter (PSO) 0.1 |
| MLR       | solver activation relu, n estimators (PSO) 1, learning_rate (PSO) 0.05 |
| LGBM      | n_estimators 100, max_delta_step 10, learning_rate (PSO) 0.01 |
|           | drop_rate 0.5, num_leaves 128, boosting_type gbdot, max_depth (PSO) 69 |
|           | activation hidden_layer_sizes (128, 256, 64), max_iter (PSO) 1500 |
|           | objective regression |

Note: The parameters with brackets “(PSO)” in the table are the target parameters for PSO optimization. Affected by factors such as computational cost and practical significance, the optimization problems are NP-hard (non-deterministic polynomial) problems, which cannot be solved by an accurate analytical algorithm. Therefore, only the approximate best values of the optimized parameters can be obtained, that is, the suboptimal values. The parameters in this table are the sub-optimal values after iterative optimization of the PSO algorithm. The meaning of the parameters of some models can be obtained by the following website:

1. XGBoost: https://xgboost.readthedocs.io/en/latest/parameter.html#general-parameters
2. RFR: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
3. SVR: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html
4. MLR: https://scikit-learn.org/stable/modules/generated/sklearn.neural-network.MLPRegressor.html
5. LGBM: https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMRegressor.html#lightgbm.LGBMRegressor

![Fig. C.1](image-url) The SHAP relationship graphs of the temperature and the three indicators most affected by the temperature distribution. (a) Temperature indicator and transit indicator. (b) Temperature indicator and driving indicator. (c) Temperature indicator and retail and recreation indicator.
recreation. However, in contrast, figure (b) shows that no matter whether the temperature is low or high, the temperature has a balanced effect on the people’s passion for driving. In general, the temperature indicator tends to influence the transit and the retail and recreation indicators, which will indirectly affect the changing pattern of load.

References
[1] Wang Z, Hong T, Li H, Ann Piette M. Predicting city-scale daily electricity consumption using data-driven models. Adv Appl Energy 2021;2:100025. https://doi.org/10.1016/j.adapen.2021.100025.
[2] Wu D, Zheng X, Yu Y, Otten D, Xia B, Singh C, et al. An open-source extendable model and correctness assurance of the 2021 texan power outage. Adv Appl Energy 2021;4:100056. https://doi.org/10.1016/j.adapen.2021.100056.
[3] Zhang H, Chen J, Yan J, Song X, Shibasaki R, Yan J. Urban power load profiles under ageing transition integrated with future EVs charging. Adv Appl Energy 2021;1:100007. https://doi.org/10.1016/j.adapen.2021.100007.
[4] Abu-Rayash A, Dincer I. Analysis of the electricity demand trends amidst the COVID-19 coronavirus pandemic. Energy Res Soc Sci 2020;68:101682. https://doi.org/10.1016/j.erss.2020.101682.
[5] Norouzi N, Zarazua de Rubens G, Choupanpiesheh S, Enevoldsen P. When pandemics impact economies and climate change: Exploring the impacts of COVID-19 on oil and electricity demand in China. Energy Res Soc Sci 2020;68:101654. https://doi.org/10.1016/j.erss.2020.101654.
[6] Chen Y, Yang W, Zhang B. Using Mobility for Electrical Load Forecasting During the COVID-19 Pandemic. ArXiv E-Prints 2020:arXiv:2006.08826.
[7] Dietrich B, Walthier J, Weigold M, Abele E. Machine learning based very short term load forecasting of machine tools. Appl Energy 2020;276:115440. https://doi.org/10.1016/j.apenergy.2020.115440.
[8] Yin J, Li X-J. Multi-temporal-spatial-scale temporal convolution network for short-term load forecasting of power systems. Appl Energy 2021;283:116328. https://doi.org/10.1016/j.apenergy.2021.116328.
[9] Ahmad T, Zhang H. Novel deep supervised ML models with feature selection approach for large-scale utilities and buildings short and medium-term load requirement forecasts. Energy 2020;209:118477. https://doi.org/10.1016/j.energy.2020.118477.
[10] Allbachti MR, El-Naggam KM. Long term electric load forecasting based on particle swarm optimization. Appl Energy 2010;87:320–6. 10.1016/j.apenergy.2009.04.024.
[11] Eapen RR, Simon Performance Analysis of Combined Similar Day and Day Ahead Short Term Electrical Load Forecasting using Sequential Hybrid Neural Networks. IETE J Res 2019;65(2):216–26. 10.1080/00171983.2017.1417749.
[12] Khoshvor A, Paouvelis EJ. Short-term scenario-based probabilistic load forecasting: A data-driven approach. Appl Energy 2019;238:1258–8. 10.1016/j.apenergy.2019.01.155.
[13] Cai M, Pipattanasomporn M, Rahuman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl Energy 2019;236:1078–88. 10.1016/j.apenergy.2018.12.042.
[14] Chen Y, Xu P, Chuy L, Wu Y, Ni Y, et al. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. Appl Energy 2017;195:659–70. https://doi.org/10.1016/j.apenergy.2017.03.034.
[15] Goswami K, Kandali AB. Electricity Demand Prediction using Data Driven Forecasting Scheme: ARIMA and SARIMA for Real-Time Load Data of Assam. 2020 Int Conf Comput Perform Eval 2020. https://doi.org/10.1109/compe49325.2020.9200031.
[16] Chen Y, Zhang D. Theory-guided deep-learning for electrical load forecasting. Adv Appl Energy 2020;266:114850. https://doi.org/10.1016/j.apenergy.2020.114850.
[17] Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Trans Knowl Data Eng 2019;31(12):2516–19. https://doi.org/10.1109/TKDE.2019.2919069.
[18] Zhang Y, Tao L. Short term power load prediction with knowledge transfer. Int Syst 2019;51(4):161–9. https://doi.org/10.1109/59.736298.
[19] Tan Z, De G, Li M, Lin H, Yang S, Huang L, et al. Combined electricity-heat-cool gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. J Clean Prod 2020;248:119252. https://doi.org/10.1016/j.jclepro.2019.119252.
[20] ZENG P, SHENG C, JIN M. A learning framework based on weighted knowledge transfer for holiday load forecasting. J Mod Power Syst Clean Energy 2019;7(2):329–39. https://doi.org/10.1007/s40565-018-0452-5.
[21] Gilsilanar M, Wang H, Sriman LMK, Ozgunen EE, Ambarchi R. Multitask Bayesian Spatiotemporal Gaussian Processes for Short-Term Load Forecasting. IEEE Trans Ind Electron 2020;67(5):5132–43. https://doi.org/10.1109/TIE.2019.2930064.
[22] Apple. Mobility Data 2020. https://www.apple.com/covid19/mobility.
[23] Google. COVID-19 Community Mobility Reports 2020. https://www.google.com/covid19/mobility/.
[24] Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions. Proc. 31st Int. Conf. Neural Inf. Process. Syst., Red Hook, NY, USA: Curran Associates Inc.; 2017, p. 4766–4777.
[25] Liu B, Nair B. Multivariate Variational Mode Decomposition. IEEE Trans Smart Grid 2018;9(1):191–8. https://doi.org/10.1109/TSG.2017.2759642.
[26] Song K-B, Ha S-K, Park J-W, Kwon D-J, Kim K-H. Hybrid load forecasting method with analysis of temperature sensitivities. IEEE Trans Power Syst 2008;23(2):889–96. https://doi.org/10.1109/TPWRS.2008.9197049.
[27] Cui L. Designing a short-term load forecasting model in the urban smart grid system. Appl Energy 2020;266:114850. https://doi.org/10.1016/j.apenergy.2020.114850.
[28] Poli R, Kennedy J, Blackwell T. Particle swarm optimization. Swarm Intel 2007;1(1):33–57. https://doi.org/10.1007/s11721-007-0002-0.