Energy Conversion and Management 297 (2023) 117715

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Regional wind-photovoltaic combined power generation forecasting based on a novel multi-task learning framework and TPA-LSTM

Yuejiang Chen, Jiang-Wen Xiao *, Yan-Wu Wang, Yuanzheng Li

School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, 430074, China
Key Laboratory of Image Processing and Intelligent Control (Huazhong University of Science and Technology), Ministry of Education, Wuhan, 430074, China

ARTICLE INFO

Keywords:
Multi-task learning
Spatio-temporal correlation
Wind-photovoltaic combined power generation forecasting

ABSTRACT

Existing renewable power generation forecasting methods mainly focus on a single energy source and fail to effectively capture the spatio-temporal correlation between different power generation resources. Furthermore, the current single-site power forecasting no longer fulfills the demands of grid dispatch. This paper introduces an innovative framework for multi-task learning and uses it to achieve regional wind-photovoltaic combined power generation forecasting. First, this paper employs Maximum Information Coefficient (MIC) to identify the crucial meteorological features affecting power generation and analyze the complementarity and correlation between wind and photovoltaic power generation. Then, an innovative multi-task learning framework is proposed that separates task-specific components and shared components, allowing each task to select adaptive information that benefits itself. Besides, this paper proposes a loss optimization strategy to balance the loss magnitude and training velocity of different tasks. In order to effectively share the coupling information among the two kinds of power generation, the proposed framework is adopted to construct the regional wind-photovoltaic combined power generation forecasting model based on Temporal Pattern Attention LSTM (TPA-LSTM) algorithm. Finally, the efficiency and superiority of the proposed method are validated through several verification and comparison case studies.

1. Introduction

Renewable energy plays a crucial role in national energy strategies as a solution to pressing challenges in energy supply and climate change. Notably, wind and photovoltaic power generation have garnered global recognition due to their renewable nature, environmental friendliness, abundant distribution of resources, and extensive utilization.

The ability to forecast wind and photovoltaic power generation in advance provides valuable insights for grid operators, energy traders, and renewable energy system planners [1]. Accurate forecasts enable efficient load balancing and support decision-making processes related to energy storage and backup generation. Furthermore, forecasting plays a important role in enhancing the economic viability and overall stability of renewable energy systems, ensuring their successful integration into the global energy mix. However, both wind and photovoltaic power are volatile and intermittent, posing significant challenges to grid security and economic stability with high renewable energy penetration [2]. Therefore, improving the accuracy and efficiency of power generation forecasting is pivotal for the power system.

Over the years, remarkable advancements have been achieved in the development of advanced forecasting techniques for wind and photovoltaic power generation. Researchers have explored various meteorological and environmental factors, historical data analysis, and real-time monitoring systems to enhance the accuracy and reliability of forecasts. However, challenges such as the complex dynamics of weather patterns, the spatial and temporal volatility of renewable resources, and the integration of diverse data sources continue to demand further research and innovation in this field.

When analyzing the key challenges of improving forecasting models for renewable energy generation, few literature consider the correlation between different generation resources. Traditional wind and photovoltaic power generation forecasting methods usually forecast each energy source independently, ignoring the mutual relationship and influence between the two [3,4]. However, the integration of wind and photovoltaic power generation through combined forecasting offers a comprehensive approach that takes into account their coupling relationship. By establishing suitable models and algorithms, accurate power generation forecasts for both energy sources can be achieved. The practical application of this combined forecasting method holds immense significance and benefits.

* Corresponding author at: School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, 430074, China.
E-mail address: jwxiao@hust.edu.cn (J.-W. Xiao).

https://doi.org/10.1016/j.enconman.2023.117715
Received 8 July 2023; Received in revised form 20 September 2023; Accepted 25 September 2023
Available online 3 October 2023

Taking into account the limitations of prior research, this study introduces a novel multi-task learning framework that utilizes the Temporal Pattern Attention LSTM (TPA-LSTM) [5] to construct a regional wind-photovoltaic combined power generation forecasting model. This model effectively captures and shares the interdependencies between the two types of power generation. This paper makes several key contributions:

(1) A novel multi-task learning framework is proposed, which introduces an innovative approach for explicitly segregating task-specific components from shared components. By leveraging the advantages of gating networks and dynamically fused representations of inputs, our framework achieves a high degree of flexibility in balancing multiple tasks. In addition, compared with the traditional multi-task learning frameworks such as MMoE [6] and PLE [7], it effectively resolves the correlation and sample dependence between conflicting tasks and improves the forecasting performance.

(2) A comprehensive forecasting model that combines regional wind and photovoltaic power generation is proposed to investigate their spatio-temporal correlation and complementarity. The proposed model can simultaneously forecast the future wind and photovoltaic power generation in the same region, which significantly improves the accuracy of regional short-term power generation forecasting compared with the separate forecasting model [8] and traditional multi-task learning frameworks include Share-Bottom [9,10], MMoE [6] and PLE [7] based on TPA-LSTM.

(3) A joint loss optimization strategy for multi-task learning model training is proposed considering the uncertainty of wind and photovoltaic power. This strategy can balance the loss magnitude and training velocity of different training difficulty tasks by learning, and adjust the loss weight in real time. Compared with the method of artificially setting the loss weight of different tasks [11,12], the proposed strategy can achieve better performance in preventing loss pulling and improving the forecasting accuracy of each task.

The rest of the paper is organized as follows. Section 2 summarizes current power forecasting methods. Section 3 analyzes the correlation between different power and its causes in detail. Section 4 introduces the combined power generation forecasting method, including the proposed multi-task learning framework, optimization strategy and modeling process. Section 5 demonstrates the application and advantages of the proposed method to regional wind-photovoltaic combined power generation forecasting. Section 6 discusses the parts of the proposed method that can be improved. Finally, Section 7 concludes the paper.

2. Literature review

Methods for renewable power forecasting can be divided into three main categories: statistical methods, traditional machine learning and deep learning algorithms. Statistical power forecasting methods involve using data to build cross-parameter correlation models to relate inputs to forecasted outputs [13]. However, they fail to consider the correlation between different input variables and the feature extraction ability is limited. Early wind and photovoltaic power generation forecasting studies have extensively employed support vector machines (SVM) [14], Gaussian process regression [15] and extreme learning machine (ELM) [16]. However, machine learning algorithms necessitate substantial training data, and models trained on fluctuating data can be susceptible to overfitting [17]. Nowadays, deep learning algorithms have gained prominence in power forecasting. These algorithms leverage large-scale multidimensional data and possess powerful mapping capabilities [18]. Yaghoubirad et al. [19] used the direct method to successfully compare Convolutional Neural Network (CNN), Long Short Term Memory Network (LSTM), CNN-LSTM and Gate Recurrent Unit (GRU) for the forecasting of wind power. Khan et al. [20] proposed a dual stream network with an attention mechanism to enhance the precision of photovoltaic power forecasting.

The deep learning algorithms mentioned above have shown promising results in forecasting wind and photovoltaic power. However, these methods mainly focus on forecasting a single form of energy, whereas the current trend is towards considering multiple generation types within renewable power clusters for practical applications [21,22]. Currently, the combined forecasting considering correlation in power systems mainly focuses on the load side [23,24], while there are few studies on the combined forecasting of renewable energy [9,10]. In reality, the weather system within a specific area exhibits inertia, and various weather elements demonstrate strong coupling and correlation. However, many existing studies tend to independently model different forecasting factors without considering the interrelationship between different renewable energy. In fact, wind and photovoltaic power clusters can be established to efficiently couple wind and photovoltaic energy, capitalizing on local power characteristics and climate complementarity [25]. Consequently, by exploring the complex correlations between the two energy sources [10,26], combining wind and photovoltaic power data can greatly improve forecasting accuracy when wind farms and photovoltaic power plants are located in the same region.

For a long time, the research on power forecasting mainly focuses on the station level, and few on regional power forecasting. However, forecasting the output power of regional cluster is an effective means to increase the power access capacity and enhance the economy and stability of power system under the condition of large-scale power access [27]. In addition, due to the smoothing effect of power generation cluster, regional forecasting has a lower error than that of single station forecasting [28]. In summary, it is of great significance to excavate the spatio-temporal correlation and complementary relationship of multiple types of renewable energy, and to construct a combined regional forecasting model to improve the power forecasting accuracy. Besides, the combined forecasting model only needs one model to obtain multiple types of forecasting results, which is more conducive to testing and maintenance than constructing one model for each energy source.

3. Correlation analysis of wind and photovoltaic power generation

Meteorological features play a significant role in power generation. Because of the complexity of atmospheric dynamics, it is difficult to quantify the relationship between power generation and meteorology, making it challenging to mathematically express their coupling. To analyze the coupling correlation between different power generation variables, the Maximum Information Coefficient (MIC) [29] is employed. MIC is a method known for its high efficiency, capable of simultaneously measuring both linear and nonlinear correlations. MIC is calculated as follows:

$$\mathrm{MIC}(x,y) = \max_{a*b<S}\left(\frac{\iint p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)}}{\log_2\min(a,b)}\right) \tag{1}$$

where S represents the sample size; a and b denote that the grid is divided into a rows and b columns; $p(x,y)$ is the joint probability function with respect to x and y; $p(x)$ and $p(y)$ are the marginal probability density functions.

The dataset for a region in a province of southern China in 2020 is used to calculate the power generation correlation. The dataset contains historical data and key meteorological features for wind and photovoltaic power generation. Each generation type contains data from three power stations, denoted as WF1-3 and PV1-3 for the wind farms and photovoltaic power plants in Fig. 1, respectively. In Fig. 1, the MIC heat maps of wind and photovoltaic power are presented. According to Fig. 1, it is evident that wind speed exerts the most significant influence on wind power generation. Additionally, wind direction and temperature also contribute significantly to the variability of wind power generation. This is because wind speed directly impacts

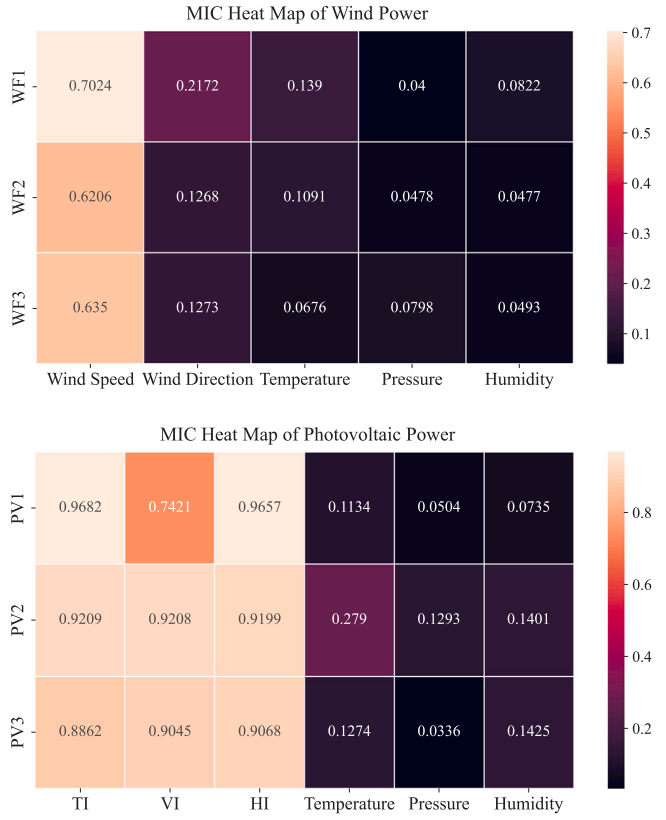

Fig. 1. MIC between power generation and meteorological features.

Table 1
Correlation coefficient of regional wind and photovoltaic power generation.

Method	Spring	Summer	Autumn	Winter
Pearson	−0.1098	−0.0873	−0.0502	−0.0927
Spearman	−0.0717	−0.0317	−0.0250	−0.0739
MIC	0.1229	0.1233	0.1179	0.1326

the amount of energy that wind turbines can harvest from the wind, and changes in wind direction and temperature can affect the efficiency of wind turbines. On the other hand, for photovoltaic power, the total irradiance (TI), vertical irradiation (VI), and horizontal irradiance (HI) are the primary features that determine the amount of power generated. Temperature and humidity are also important features in photovoltaic power generation as they affect the efficiency of energy conversion.

In the upcoming analysis, we will investigate the correlation between the two types of regional power generation across various seasons. Table 1 shows the correlation values calculated by different methods, where Pearson [30] is used to characterize the linear correlation and Spearman [31] is used to characterize the monotonic correlation. It is clear that wind and photovoltaic power generation show a weak negative monotonic correlation in all four seasons, which reflects the complementarity between them. In addition, the MIC correlation coefficients of the two types of power generation are greater than 0.1, which reflects the complex nonlinear relationship between them. The coupling relationship can be further explained by examining the impact of meteorological features. The following are the explanations for the correlation:

(1) Wind speed and irradiance have a certain correlation in meteorology. Wind is caused by a pressure gradient that arises from disparities in heat distribution among regions, and the intensity of sunlight exposure can impact the amount of heat that the land surface absorbs, affecting the strength of the wind. Thus, changes in the irradiance level can indirectly impact wind power generation.

(2) Meteorological factors, such as temperature, air pressure, and humidity, play a crucial role in influencing the generation of both wind and photovoltaic power. These features are highly interdependent and can affect both forms of power generation simultaneously due to the inertial nature of the weather system in a region.

In summary, a complex nonlinear correlation between wind and photovoltaic power generation has been confirmed, which provides theoretical support for combined power generation forecasting.

4. Combined power generation forecasting method

4.1. The novel multi-task learning framework

Multi-task learning enhances the learning process of individual sub-tasks by leveraging the shared information, resulting in improved accuracy and efficiency. Fig. 2 illustrates the distinction between the training processes of different kinds of learning methods. The effectiveness of multi-task learning in enhancing forecasting performance stems from both the inherent capabilities of the model and the strategic optimization of loss functions employed.

Previous multi-task learning frameworks, including hard parameter sharing, cross-stitch [32], sluice network [33] and Multi-gate Mixture-of-experts (MMOE) [6], have encountered challenges such as negative transfer phenomenon and seesaw phenomenon in loosely correlated and complex correlated tasks. While Progressive Layered Extraction (PLE) [7] partially resolves the seesaw problem, it does not completely disentangle the task-specific information from the shared information in regression task. Drawing inspiration from PLE, an innovative multi-task learning framework for combined power generation forecasting is presented in Fig. 3. This framework aims to overcome the aforementioned limitations and achieve better separation of task-specific and shared information. The proposed framework comprises three main components: Input Module, Extraction Module and Output Module.

1. Input Module: The input module is used to integrate different task inputs and unify the dimensions. The input is the input data for each task, and the output is the adjusted task-specific information and shared information. The output of input module can be calculated as:

$$\tilde{i}^a = W_i^a i^a + b^a \tag{2}$$

$$\tilde{i}^s = W_i^s [i^a, i^b] + b^s \tag{3}$$

$$\tilde{i}^b = W_i^b i^b + b^b \tag{4}$$

where i^a and i^b are the input data for tasks A and B; W_i^a, W_i^s, W_i^b are weight matrices; b^a, b^s, b^b are bias matrices; $\tilde{i}^a$, $\tilde{i}^s$, $\tilde{i}^b$ are outputs after linear transformation with the same dimension.

2. Extraction Module: The role of the extraction module is to extract task-specific information and shared information. An extraction module can include one or more extraction networks. For one extraction network, there are gating networks at the top connected with the corresponding shared and task-specific expert modules at the bottom. Each expert module consists of multiple experts, where each expert is a feature extraction network. Different experts in the extraction network are responsible for learning different information. The formulation for the output of task k in a single extraction network is as follows:

$$g^k(x_k) = w^k(x_k)[S^k, S^s]^T \tag{5}$$

$$w^k(x_k) = Softmax(W_g^k x_k) \tag{6}$$

where x_k is the input of task k's experts in the extraction network, x_k is equal to $\tilde{i}^k$ when the extraction module is the first module connected after the input module; $w^k(x_k)$ is a weight

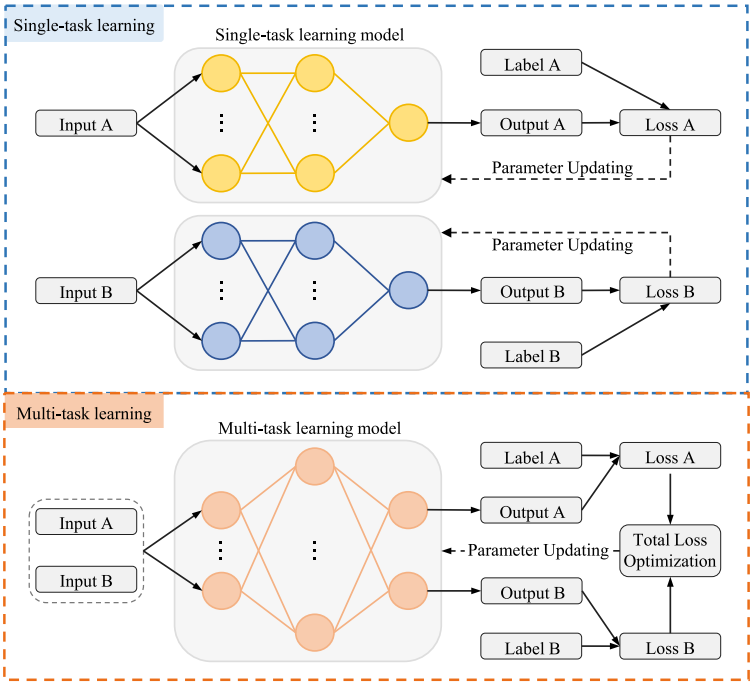

Fig. 2. The training process of single-task learning and multi-task learning.

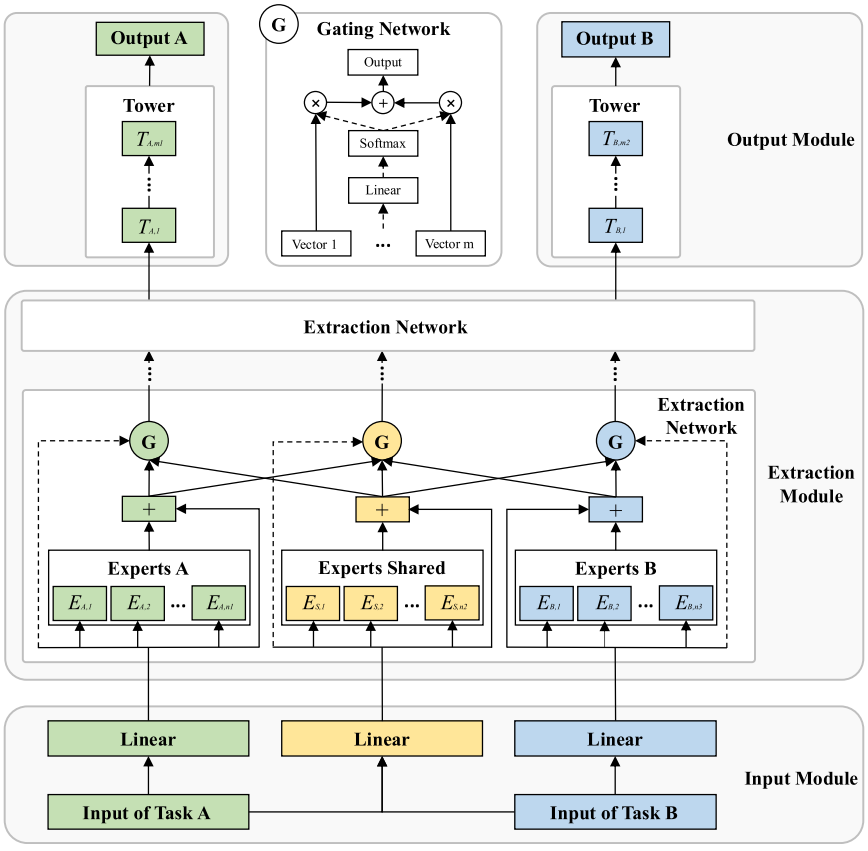

Fig. 3. The novel multi-task learning framework.

function to calculate the weight vector; W_g^k is the parameter matrix; S^k and S^s are the selection matrices for task k and the shared part. These components can be shown as:

$$S^k = \left[E^T_{(k,1)}, E^T_{(k,2)}, \cdots, E^T_{(k,n_k)}, x^T_k\right] \tag{7}$$

$$S^s = \left[E^T_{(s,1)}, E^T_{(s,2)}, \cdots, E^T_{(s,n_s)}, x^T_s\right] \tag{8}$$

where x_s is the input of shared experts in the extraction network, x_s is equal to $\tilde{l}^s$ when the extraction module is the first module

connected after the input module; E represents the output of the expert network; From (7)–(8), we can know that the selected matrices includes not only the experts E but also the input x, which is connected by the residual structure. Experts here can select different networks according to different application scenarios.

3. Output Module: The output module is used to output the forecasted values for each task. There is a dedicated tower network for each task:

$$y^k(x_k) = t^k(g^{k,N}(x_k)) \tag{9}$$

where N is the number of extraction networks; t^k is the tower network of task k; $g^{k,N}(x_k)$ represents the output of task k in the last extraction network. The tower network here varies according to the application scenario. In the time series forecasting task, the tower network can choose models such as LSTM, GRU.

The proposed framework improves the efficiency of compositional representation learning and cross-task information routing by explicitly separating task-specific components and shared components. Furthermore, the proposed multi-task learning framework incorporates gating networks, which enable dynamic fusion of representations based on the input data. This approach allows for a flexible balance between tasks and enhances the framework's ability to handle task conflicts and sample dependencies effectively. Compared to PLE, the proposed framework offers the following improvements:

(1) The input module of the proposed framework divides the overall input data into task-specific input data and combined input data, which are then respectively fed to task-specific and shared experts. This approach allows task-specific experts to learn only the information relevant to their task without interference from other tasks, while shared experts focus on learning shared information between different tasks.

(2) The framework uses residual structure in the extraction module, and the input of the expert is also used as an expert to participate in the decision. This method prevents the influence of the degradation of the expert module on the final forecasting results, and allows the model to adaptively select the information beneficial to the forecasting by means of weighting.

4.2. Optimization strategy

The effectiveness of multi-task learning models heavily relies on the weighting scheme between different tasks' loss functions. At present, the commonly used method is to weight the loss of each task to obtain the total loss [11,12], as shown in (10).

$$L = \sum_k \omega_k L_k \tag{10}$$

where L represents the total loss function; ω_k and L_k denote the loss weight and loss function of task k. Manually adjusting the weight of each loss can be resource-intensive and may not result in an optimal weight for each task.

In recent years, several multi-objective optimization strategies have been proposed at two levels: balancing loss magnitude and adjusting training velocity. These strategies include uncertainty weight [34], gradient normalization [35], and dynamic weight averaging [36]. Existing optimization methods have not been successful in providing a convenient and efficient approach to simultaneously balance the magnitude of loss and training velocity. To overcome this limitation, this paper introduces a novel optimization strategy that combines uncertainty weight and dynamic weight averaging methods. This strategy comprehensively considers the loss magnitude and training velocity based on the homoscedastic uncertainty of different tasks. The final loss function $L(t)$ is obtained as follows:

$$r_k(t-1) = \frac{L_k(t-1)}{L_k(t-2)} \tag{11}$$

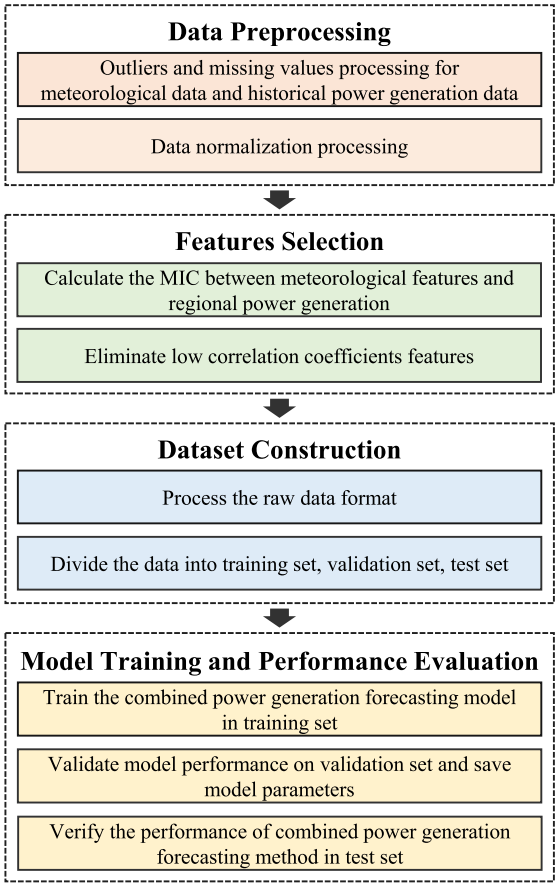

Fig. 4. The process of combined power generation forecasting.

$$\omega_k(t) = \frac{n\exp\left(r_k(t-1)/T\right)}{\sum_{k=1}^{n}\exp\left(r_k(t-1)/T\right)} \tag{12}$$

$$L(t) = \sum_{k=1}^{n}\left(\frac{\omega_k(t)}{2\sigma_k^2}L_k(t) + \log\sigma_k\right) \tag{13}$$

where t is the current time point; n represents the number of tasks; $\omega_k(t)$ is used to regulate the training velocity of task k; T is a constant; The parameter σ_k represents the homoscedastic uncertainty of task k and needs to be trained during the optimization process. A smaller σ_k means a lower noise for the task k, and a higher weight assigns to the loss of task k. After obtaining the total loss, the Adam optimizer [37] can be applied to optimize the model parameters.

4.3. Modeling process

Building upon the aforementioned analysis, the combined power generation forecasting model is constructed. The flowchart of the regional wind and photovoltaic combined power generation forecasting, taking into account the coupling relationship, is depicted in Fig. 4. The model follows specific steps, which are outlined below:

(1) Data preprocessing: The first step involves searching for outliers and missing values in the meteorological and historical power data. These anomalies are then filled with data from the previous moment to ensure continuity. Subsequently, the data is normalized to achieve a consistent data scale, which helps accelerate the training velocity.

(2) Features selection: The MIC method is utilized to calculate correlation coefficients between meteorological features and regional power generation. Features with low correlation coefficients are then eliminated from the analysis.

(3) Dataset construction: To handle time series data, a sliding window approach is employed to preprocess the raw data, transforming it into a format suitable for training. Subsequently, the data for each month is partitioned into training set, validation set, and test set.

(4) Model training and performance evaluation: The regional power generation forecasting model is trained and evaluated. The model's best-performing parameters on the validation set are saved for making forecasts on the test set. Comparative simulations are subsequently conducted to assess the accuracy of the proposed method on the test set.

To compare the performance of power generation forecasting, the evaluation indicators used are R-squared (R^2) and Root Mean Squared Error (RMSE). These indicators are employed to assess and evaluate the accuracy of the forecasting results [38].

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i' - y_i\right)^2}{\sum_{i=1}^{n} \left(\bar{y} - y_i\right)^2} \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(y_i' - y_i\right)^2} \tag{15}$$

where n represents the number of samples, y_i denotes the true data, and y_i' represents the forecasted data of power generation. R^2 measures the goodness of fit between the forecasted and true data, with higher values indicating a more accurate fit. On the other hand, RMSE calculates the error between the forecasted and true data, with lower values indicating lower forecasting deviation. These indicators possess distinct characteristics that enable a comprehensive evaluation of the forecast results.

5. Case studies

This paper focuses on analyzing historical data from wind and photovoltaic power stations in a specific region of a southern province in China during the year 2020. The dataset comprises 366 days of data, with a time interval of 15 minutes, resulting in a total of 35,136 data points. This dataset includes historical records of wind and photovoltaic power generation as well as essential meteorological features, making it suitable for conducting power generation forecasting simulations. The regional power generation is determined by aggregating the power output from similar types of power stations, while the meteorological data from each station is utilized as the meteorological features for the regional power generation. The target of this research is to forecast power generation for the subsequent 15 minutes based on the preceding 2 hours of data. All simulations are carried out using the Python programming language within the Jupyter Notebook platform. The PyTorch deep learning library is employed to implement the power generation forecasting algorithm and training process.

5.1. Data preprocessing

In this study, the box plot method [39] is used to identify outliers, which are then treated as missing data. The missing data are filled with the data from the previous moment to ensure continuity of the time series data. The box plot method, in comparison to other commonly used outlier detection methods like 3-sigma [40] and z-score [41], offers the advantage of not requiring the data to follow a normal distribution. There are no missing data in our dataset, but the outliers values of power and meteorological data account for 1.67% and 4.16% of the wind and photovoltaic power generation, respectively.

The data in this paper include meteorological and power generation data. Different features may have different dimensions and magnitudes. Without normalization, some indicators may be overlooked, potentially affecting the accuracy of data analysis. To address this, the max–min normalization method is used in this paper to scale the data to the range of 0–1.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{16}$$

Table 2
MIC Correlation coefficient of meteorological features and regional wind power.

	WS	WD	T	P	H
WF1	**0.241**	**0.106**	0.071	0.039	0.063
WF2	**0.512**	**0.123**	**0.102**	0.043	0.042
WF3	**0.123**	0.054	0.077	**0.110**	0.042

Table 3
MIC Correlation coefficient of meteorological features and regional photovoltaic power.

	TI	VI	HI	T	P	H
PV1	**0.905**	**0.675**	**0.902**	**0.139**	**0.101**	0.082
PV2	**0.719**	**0.719**	**0.718**	**0.133**	0.090	0.090
PV3	**0.756**	**0.769**	**0.767**	**0.134**	0.039	**0.190**

where x and x' represent the original and normalized series of the variables; In a series x, $\min(x)$ represents the minimum value, while $\max(x)$ represents the maximum value.

5.2. Features selection

The regional wind power cluster contains three wind power stations. In addition to the annual power generation data of each wind power station, the historical dataset also encompasses five meteorological features for each station. These features include wind speed (WS), wind direction (WD), temperature (T), pressure (P), and humidity (H). For wind direction, we use sin function to convert the angle value. Similar to regional wind power cluster, the regional photovoltaic power cluster also contains three photovoltaic power stations. The historical data consists of the annual power generation data and six meteorological features for each photovoltaic power station. These meteorological features include total irradiance, vertical irradiation, horizontal irradiance, temperature, pressure, and humidity.

To improve efficiency and save resources, identifying significant meteorological features that impact regional power generation is crucial. The MIC method is used to determine the correlation between meteorological features and regional power generation, and only those with correlation coefficients greater than 0.1 are selected as features for model training. The selected meteorological features for model training can be found in Tables 2–3, which show the correlation coefficients between the meteorological features of each power station and the regional power generation.

5.3. Dataset construction

In regression tasks, it is essential to partition the raw data into feature data (x) and target data (y) prior to training. Considering the regional historical data of power generation and the selected key meteorological features mentioned earlier, the forecasting model can be formulated as follows:

$$w(t+1) = f\left(w(t), \ldots, w(t-T), m(t), \ldots, m(t-T)\right) \tag{17}$$

where $w(t+1)$ is the forecasted power generation and T represents the sequence length of the historical data. As input data to the model, $w(t), \ldots, w(t-T)$ and $m(t), \ldots, m(t-T)$ are historical power generation and meteorological features, respectively. The method of using sliding window processing raw data to obtain feature data x and target data y for training is shown in Fig. 5.

The last step in dataset construction involves dividing the feature data and target data into training set, validation set and test set. In this study, the data for each month of the year is divided by following an 8:1:1 ratio. Furthermore, the data is aggregated based on seasons to improve the reliability and credibility of the simulations. The distribution of the resulting dataset can be found in Table 4.

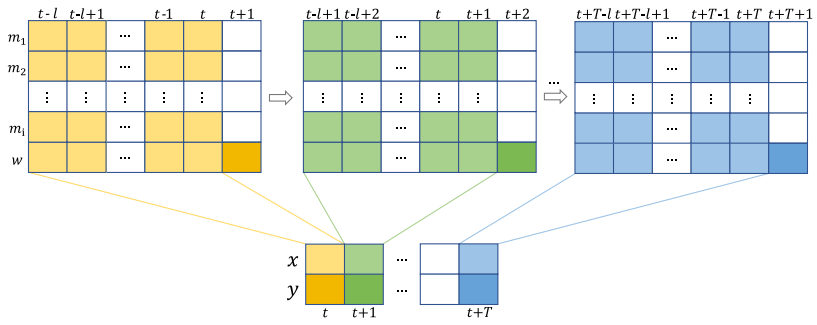

Fig. 5. Processing data by sliding window.

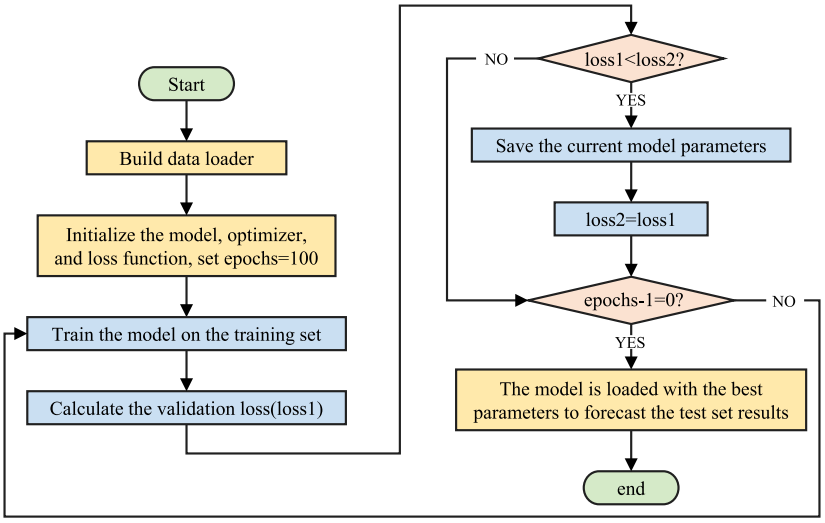

Fig. 6. Flow chart of training process.

Table 4
Dataset division in different seasons.

	Spring	Summer	Autumn	Winter	Total
Training set	7045	7045	6968	6968	28206
Validation set	881	881	871	872	3505
Test set	882	882	873	872	3509

5.4. Performance of regional power generation forecasting based on TPA-LSTM

In this paper, TPA-LSTM [5] is selected as the tower network for combined forecasting. In order to verify the forecasting performance of TPA-LSTM, this subsection compares it with five powerful and widely used forecasting networks, namely CNN, LSTM, Bi-directional Long Short-Term Memory network (BILSTM) [23] and GRU.

To address overfitting, we incorporate the recurrent neural network with a single hidden layer and a dropout rate of 0.2. L2 regularization [42] is applied using a regularization parameter of 0.01. The Rectified Linear Unit (ReLU) [43] activation function is used, and the loss function employed is the root mean square error. We utilize the Adam optimizer to update the model parameters. The remaining parameters are explored within the specified range in Table 5, and the combination that exhibits the best performance on the test set is selected as the optimal outcome. The training process is depicted in Fig. 6. Each model undergoes five training iterations, and the average value is taken as the final result.

In this part, in order to compare the performance of the five networks, we use the five networks to forecast the wind and photovoltaic power separately without using the combined forecasting model. The average power generation forecasting results of the four seasons are shown in Table 6. Comparing the forecasting effects of wind and photovoltaic power generation, it is evident that the fitting effect and forecasting error of photovoltaic power generation are better than that of wind power generation, which indicates that stable and periodic data can achieve better forecasting performance.

Among the five networks, CNN performs the worst in both wind and photovoltaic power forecasting. LSTM and BILSTM have similar forecasting performance, but GRU has a small decrease compared to them. Among the different models evaluated, TPA-LSTM demonstrates

Table 5
Parameter range of forecasting model.

	Learning rate	Hidden neurons	Batch size
Scope	[0.1, 0.01, 0.001]	[16, 32, 48]	[32, 512]

Table 6
Performance comparison of different baseline models.

Power type	Method	R^2 (%)	RMSE (MW)
Wind power	CNN	96.877	10.368
	LSTM	97.272	9.765
	BILSTM	97.275	9.751
	GRU	97.280	9.738
	TPA-LSTM	**97.287**	**9.727**
Photovoltaic power	CNN	98.454	7.345
	LSTM	98.696	6.795
	BILSTM	98.692	6.823
	GRU	98.683	6.897
	TPA-LSTM	**98.752**	**6.633**

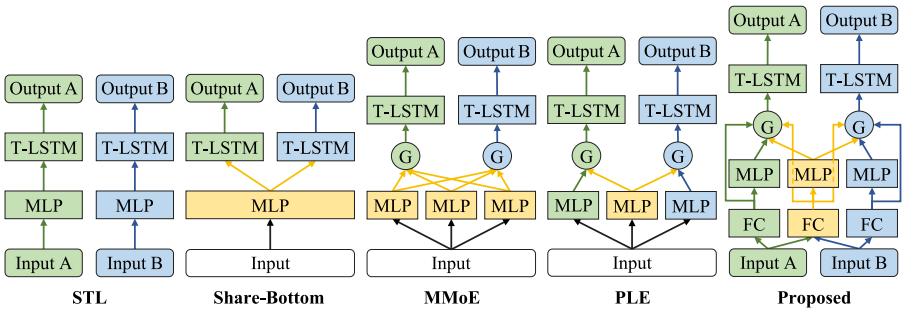

Fig. 7. Model structures for comparison.

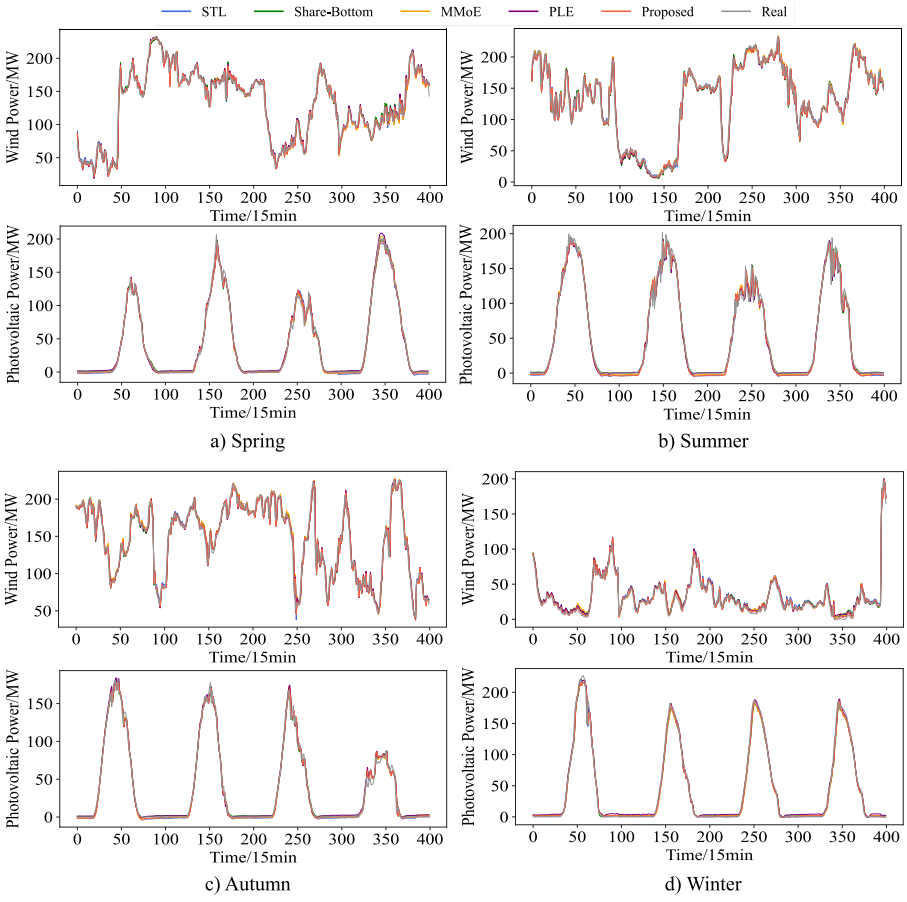

Fig. 8. Forecasting curves of wind and photovoltaic power in different seasons.

superior performance in forecasting both wind and photovoltaic power. This network achieves the best results with a configuration consisting of 32 hidden neurons, a learning rate of 0.001, and a batch size of 512. For wind power forecasting, the improvement of TPA-LSTM is very small, with R^2 increasing by 0.007% and RMSE decreasing by 0.011 MW. In the photovoltaic power forecasting, TPA-LSTM has a very significant improvement, with R^2 increases by 0.056% and RMSE decreases by 0.162 MW. Due to its outstanding performance in regional power generation forecasting, TPA-LSTM is chosen as the tower network for combined forecasting.

5.5. Performance of combined power forecasting based on the proposed multi-task learning framework and TPA-LSTM

The purpose of this section is to assess the effectiveness of the proposed multi-task learning framework in capturing shared information between wind and photovoltaic power generation and enhancing forecasting performance. To achieve this, we compare the proposed model with a single-task learning model [8]. We also assess the effectiveness of commonly used multi-task learning models, including Share-Bottom [9,10], MMoE [6], and PLE [7]. Each model structure is shown in Fig. 7, where T-LSTM represents the TPA-LSTM network, and the network configuration and hyperparameters have been determined in the previous subsection. A two-layer Multilayer Perceptron (MLP) is set as an expert to learn and extract relevant information, where each layer contains 16 neurons. For the proposed model, the Fully Connection (FC) layer with 16 neurons is used to adjust the dimension of the input. The proposed loss optimization strategy is used to balance the loss magnitude and training velocity of multi-task learning. The training configurations such as learning rate and optimizer are exactly the same as in the previous section.

The proposed model is evaluated by obtaining the power forecasting results for different seasons. Fig. 8 shows the forecasting results of

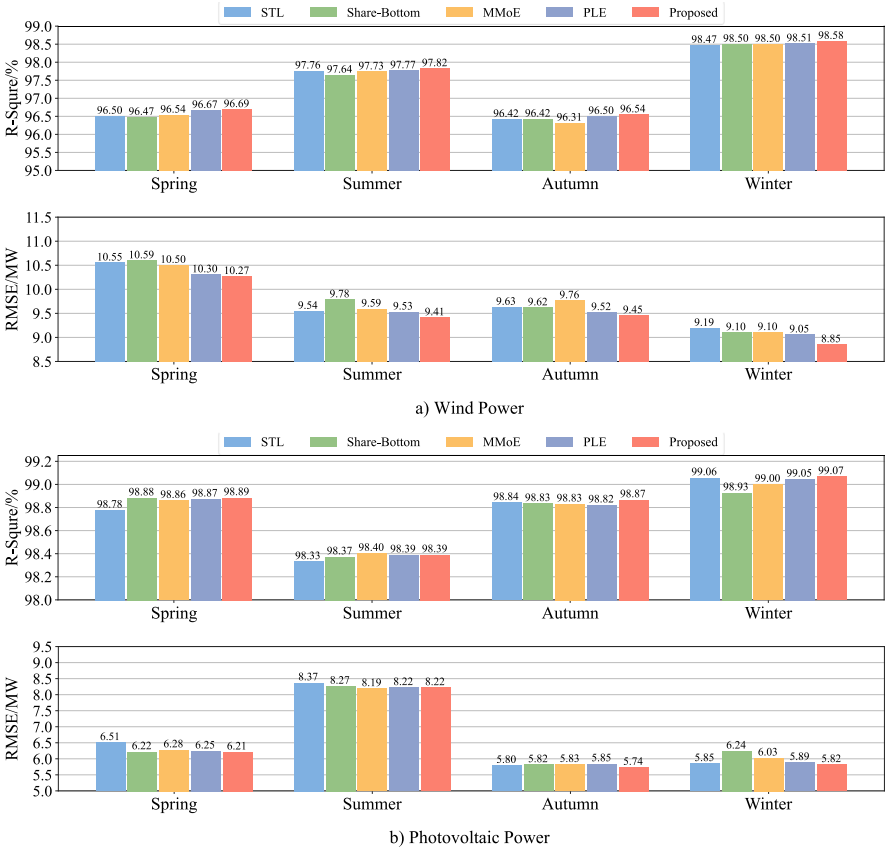

Fig. 9. Forecasting result of wind and photovoltaic power in different seasons.

Table 7
Power forecasting results of different models.

Method	Wind power		Photovoltaic power	
	R^2 (%)	RMSE (MW)	R^2 (%)	RMSE (MW)
STL	97.287	9.727	98.752	6.633
Share-Bottom	97.260	9.773	98.754	6.637
MMoE	97.271	9.739	98.774	6.581
PLE	97.360	9.601	98.783	6.551
Proposed	**97.408**	**9.495**	**98.803**	**6.498**

wind and photovoltaic power generation for 400 time points (100 h) in each season. The proposed model exhibited better fitting performance than other methods with curves that are closer to the real curve. Table 7 presents the average goodness-of-fit and forecast error for the whole year. Compared with the single-task based generation forecasting, the proposed multi-task forecasting achieved an increase in R^2 from 97.287% to 97.408% and a decrease in RMSE from 9.727 MW to 9.495 MW for wind power generation. For photovoltaic power generation, the proposed multi-task forecasting achieved an increase in R^2 from 98.752% to 98.803% and a decrease in RMSE from 6.633 MW to 6.498 MW. Besides, it is observed that the commonly used multi-task learning models such as Share-Bottom and MMoE are found to have negative transfer or seesaw phenomenon compared with the single-task learning model. Compared with PLE, the proposed model has 0.048% and 0.020% performance improvement in R^2, as well as 0.106 MW and 0.053 MW performance improvement in RMSE. In general, compared with the single forecasting model and other combined forecasting models, our proposed model can simultaneously improve the forecasting accuracy of the two types of power generation.

The goodness-of-fit and forecasting error of each model for different seasons are displayed in Fig. 9. Wind power has the worst forecasting performance in spring and the best performance in winter. For photovoltaic power, it has the worst forecasting performance in summer and similar performance in the other three seasons. This is because the fluctuation of photovoltaic power is the strongest in summer, thus increasing the difficulty of forecasting. Even though our model is not the best performing model in summer, our model ranks second among the five models and is almost the same as the best model in all indicators. Overall, the proposed multi-task learning model performs the best in most seasons and evaluation indicators, accounting for 87.5% of cases, which demonstrates the robustness of our model.

5.6. Verify the effectiveness of the improvements of the proposed framework

In Section 4.1, we have mentioned that the proposed multi-task learning framework has two improvements over PLE. One is that the input module of the proposed framework divides the overall input data into task-specific input data and combined input data, the other is that the framework uses residual structure in the extraction module. To assess the efficacy of the improvements in the proposed framework, we will capture and average the output of the gating network. This approach enables us to analyze the impact and performance of the improvements within the proposed model. A pseudo-expert is added to the task-specific expert module in this section, and the output of the pseudo-expert is constantly a random number between 0 and 1 to verify that the gating network can adaptively select information favorable to the result. The extraction layer with the pseudo-experts added is shown in Fig. 10.

As shown in Fig. 10, there are two gating networks that can set weight coefficients for different experts. The outputs of the two gating networks are shown in Fig. 11. It can be clearly seen that the weights of the two pseudo-experts are very low, which are 0.033 and 0.038 respectively, indicating that the gating network can effectively select

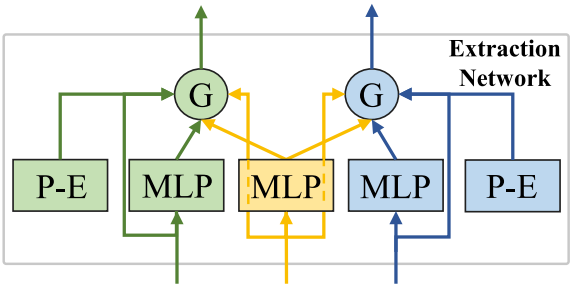

Fig. 10. The extraction layer after adding pseudo-experts.

out useful information. For the two types of power generation, the weights of the shared part are 0.216 and 0.195, showing that the shared information is helpful to improve the forecasting accuracy. In addition, the input of experts accounts for 70.4% and 52.1% by summing the weight coefficients of the expert-inputs shown in Fig. 11, which shows the effective of the residual structure proposed in this paper. In summary, our proposed structure can well separate task-specific information and shared information, and let each module play a role in different forecasting task through the two improvements.

5.7. Performance of the proposed optimization strategy for multi-task learning

To compare the effectiveness of the proposed optimization strategy with the traditional optimization method [11,12] shown in (10), we set different loss weights and conduct simulations on the proposed model. The training loss curves for the first 25 epochs with different methods are shown in Fig. 12, Loss 1 means the training loss of wind power forecasting task (task 1) and Loss 2 means the training loss of photovoltaic power forecasting task (task 2). It can be clearly seen that when the traditional manual weighting method is adopted, there are different degrees of gap in the loss of the two tasks and they are all at a large level. When the proposed dynamic weighting method is used, the training losses of the two tasks are very close, with the maximum gap of 2.97×10^{-2} and the average gap of 0.56×10^{-3}, indicating that the proposed optimization strategy can effectively balance the loss magnitude and training velocity of different tasks.

Table 8
Lowest validation loss with different loss weights.

Val. Loss/ $\times 10^{-3}$	0.1	0.3	0.5	0.7	0.9	Proposed
Loss 1	1.66	1.35	1.07	1.45	1.36	**0.97**

Table 8 presents the lowest validation loss in the first 25 epochs with different loss weights. We can see that the proposed optimization strategy has the lowest validation loss, which is 0.97×10^{-3} and 1.22×10^{-3} respectively. In particular, it has the largest improvement on task 2, with an average improvement of 54.7% compared with the traditional weighted method. In summary, our proposed optimization strategy can achieve real-time and adaptive adjustment of loss weights and show excellent performance.

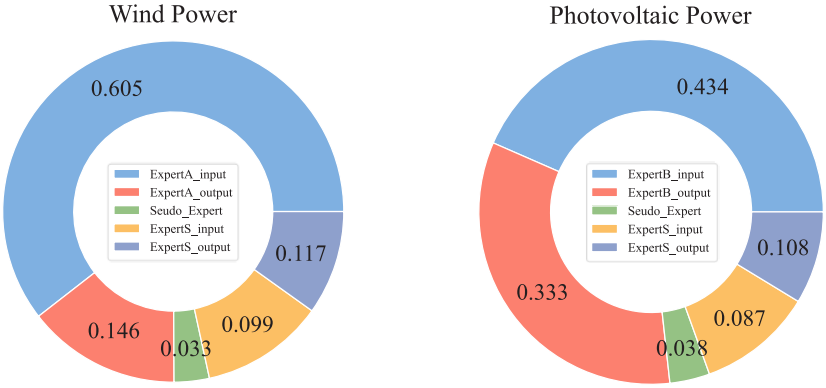

Fig. 11. The weight coefficient distribution of the experts.

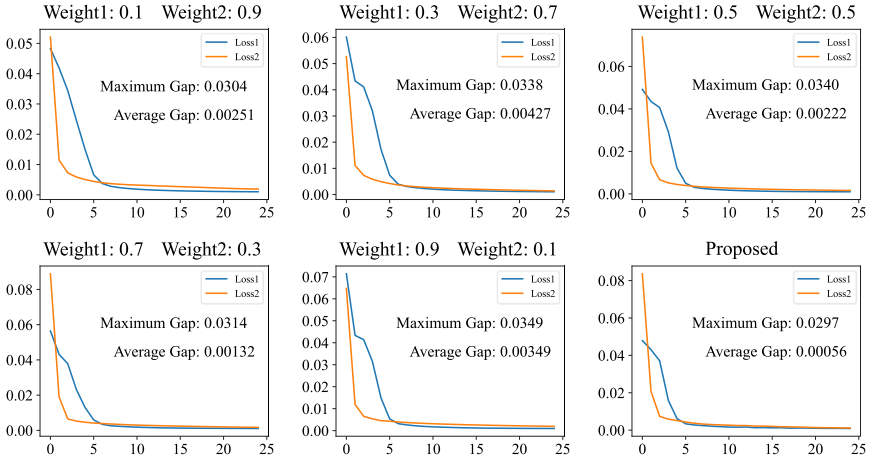

Fig. 12. Training loss with different loss weights.

6. Discussion

This paper presents a novel multi-task learning framework and a new optimization strategy for combined wind-photovoltaic power generation forecasting. Verification and comparison simulation results demonstrate that incorporating the spatio-temporal correlation between two types of power generation can significantly enhance the forecasting accuracy. While this paper offers valuable research contributions, there are still areas that can be further improved. For instance, in the proposed multi-task learning framework, more advanced feature extraction models or an increased number of extraction modules can be explored to extract deeper information. Additionally, incorporating the uncertainty of power generation, probabilistic forecasting techniques can be employed to provide more comprehensive forecasting information, replacing deterministic forecasting for the model's output. These improvements would enhance the overall forecasting capabilities and provide valuable insights.

7. Conclusion

This paper introduces a wind-photovoltaic combined power generation forecasting model based on multi-task learning. The proposed model takes into account the spatio-temporal correlation between wind and photovoltaic power. The MIC method is firstly used to analyze the correlation between wind and photovoltaic power. Then we propose a novel multi-task learning framework and loss optimization strategy. The proposed framework is used to construct a regional wind and photovoltaic combined power generation forecasting model based on the TPA-LSTM algorithm, and the parameters are updated using the proposed optimization strategy. The proposed method ensures that each task can adaptively learn the coupling information of wind and photovoltaic without loss of accuracy. By conducting extensive and meticulous simulations using authentic datasets, the efficacy of the proposed integrated forecasting approach has been substantiated across various seasons. The verification and comparison results unequivocally demonstrate that the suggested combined power forecasting method substantially enhances the precision of power generation forecasting. Consequently, it holds significant practical significance within the domain of renewable energy power generation systems.

CRediT authorship contribution statement

Yuejiang Chen: Methodology, Data curation, Writing – original draft. **Jiang-Wen Xiao:** Conceptualization, Writing – review & editing. **Yan-Wu Wang:** Supervision, Writing – review & editing, Funding acquisition. **Yuanzheng Li:** Methodology.

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.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work is support by the National Natural Science Foundation of China under Grants 62233006, 62173152 and the Fundamental Research Funds for the Central Universities, China (HUST: 2022JYCXJJ005).

References

[1] Liao Q, Cao D, Chen Z, et al. Probabilistic wind power forecasting for newly-built wind farms based on multi-task Gaussian process method. Renew Energy 2023;217:119054.
[2] Notton G, Nivet ML, Voyant C, et al. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. Renew Sustain Energy Rev 2018;87:96–105.
[3] Sun S, Liu Y, Li Q, et al. Short-term multi-step wind power forecasting based on spatio-temporal correlations and transformer neural networks. Energy Convers Manage 2023;283:116916.
[4] Sarmas E, Spiliotis E, Stamatopoulos E, et al. Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent long short-term memory models. Renew Energy 2023;216:118997.
[5] Shih SY, Sun FK, Lee H. Temporal pattern attention for multivariate time series forecasting. Mach Learn 2019;108:1421–41.
[6] Ma J, Zhao Z, Yi X, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018, p. 1930–9.
[7] Tang H, Liu J, Zhao M, et al. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In: Proceedings of the 14th ACM conference on recommender systems. 2020, p. 269–78.
[8] Huang B, Liang Y, Qiu X. Wind power forecasting using attention-based recurrent neural networks: A comparative study. IEEE Access 2021;9:40432–44.
[9] Sanjari MJ, Gooi HB, Nair NKC. Power generation forecast of hybrid PV-wind system. IEEE Trans Sustain Energy 2019;11(2):703–12.
[10] Zhang Y, Yan J, Han S, et al. Joint forecasting of regional wind and solar power based on attention neural network. In: 2022 IEEE 5th international electrical and energy conference. IEEE; 2022, p. 4165–9.
[11] Kokkinos I. Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 6129–38.
[12] Teichmann M, Weber M, Zoellner M, et al. Multinet: Real-time joint semantic reasoning for autonomous driving. In: 2018 IEEE intelligent vehicles symposium. IEEE; 2018, p. 1013–20.
[13] Mansoor M, Mirza AF, Usman M, et al. Hybrid forecasting models for wind-PV systems in diverse geographical locations: Performance and power potential analysis. Energy Convers Manage 2023;287:117080.
[14] Yuan X, Chen C, Yuan Y, et al. Short-term wind power prediction based on LSSVM-GSA model. Energy Convers Manage 2015;101:393–401.
[15] Kou P, Liang D, Gao F, et al. Probabilistic wind power forecasting with online model selection and warped Gaussian process. Energy Convers Manage 2014;84:649–63.
[16] Yin H, Dong Z, Chen Y, et al. An effective secondary decomposition approach for wind power forecasting using extreme learning machine trained by crisscross optimization. Energy Convers Manage 2017;150:108–21.
[17] Xu Q, He D, Zhang N, et al. A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. IEEE Trans Sustain Energy 2015;6(4):1283–91.
[18] Ahmed R, Sreeram V, Togneri R, et al. Computationally expedient photovoltaic power forecasting: a lstm ensemble method augmented with adaptive weighting and data segmentation technique. Energy Convers Manage 2022;258:115563.
[19] Yaghoubirad M, Azizi N, Farajollahi M, et al. Deep learning-based multistep ahead wind speed and power generation forecasting using direct method. Energy Convers Manage 2023;258:115563.
[20] Khan ZA, Hussain T, Baik SW. Dual stream network with attention mechanism for photovoltaic power forecasting. Appl Energy 2023;338:120916.
[21] Lu L, Yuan W, Su C, et al. Optimization model for the short-term joint operation of a grid-connected wind-photovoltaic-hydro hybrid energy system with cascade hydropower plants. Energy Convers Manage 2021;236:114055.
[22] Ren G, Liu J, Wan J, et al. Investigating the complementarity characteristics of wind and solar power for load matching based on the typical load demand in China. IEEE Trans Sustain Energy 2021;13(2):778–90.
[23] Guo Y, Li Y, Qiao X, et al. Bilstm multitask learning-based combined load forecasting considering the loads coupling relationship for multienergy system. IEEE Trans Smart Grid 2022;13(5):3481–92.
[24] Wang C, Wang Y, Ding Z, et al. A transformer-based method of multi-energy load forecasting in integrated energy system. IEEE Trans Smart Grid 2022;13(4):2703–14.
[25] Ravada BR, Tummuru NR, Ande BNL. Photovoltaic-wind and hybrid energy storage integrated multi-source converter configuration for DC microgrid applications. IEEE Trans Sustain Energy 2020;12(1):83–91.
[26] Han S, Qiao Y, Yan J, et al. Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. Appl Energy 2019;239:181–91.
[27] Lobo MG, Sanchez I. Regional wind power forecasting based on smoothing techniques, with application to the Spanish peninsular system. IEEE Trans Power Syst 2012;27(4):1990–7.

[28] Yang M, Zhang L, Cui Y, et al. Investigating the wind power smoothing effect using set pair analysis. IEEE Trans Sustain Energy 2019;11(3):1161–72.
[29] W Pan. Feature selection algorithm based on maximum information coefficient. In: 2021 IEEE 5th advanced information technology, electronic and automation control conference. IEEE; 2021, p. 2600–3.
[30] Guo X, Zhan Y, Zheng D, et al. Research on short-term forecasting method of photovoltaic power generation based on clustering SO-GRU method. Energy Rep 2023;9:786–93.
[31] Huang C, Yang M. Memory long and short term time series network for ultra-short-term photovoltaic power forecasting. Energy 2023;127961.
[32] Misra I, Shrivastava A, Gupta A, et al. Ross-stitch networks for multi-task learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, p. 3994–4003.
[33] Ruder S, Bingel J, Augenstein I, et al. Sluice networks: Learning what to share between loosely related tasks. 2017, 24003, arXiv preprint arXiv:1705.08142.
[34] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7482–91.
[35] Chen Z, Badrinarayanan V, Lee CY, et al. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In: International conference on machine learning. PMLR; 2018, p. 794–803.
[36] Liu S, Johns E, Davison AJ. End-to-end multi-task learning with attention. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, p. 1871–80.
[37] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.
[38] Liu ZF, Li LL, Tseng ML, et al. Prediction short-term photovoltaic power using improved chicken swarm optimizer-extreme learning machine model. J Clean Prod 2020;248:119272.
[39] Thirumalai C, Kanimozhi R, Vaishnavi B. Data analysis using box plot on electricity consumption. In: 2017 International conference of electronics, communication and aerospace technology, vol. 2, IEEE; 2017, p. 598–600.
[40] Liang R, Chang X, Jia P, et al. Mine gas concentration forecasting model based on an optimized BiGRU network. ACS Omega 2020;5(44):28579–86.
[41] Aggarwal V, Gupta V, Singh P, et al. Detection of spatial outlier by using improved Z-score test. In: 2019 3rd International conference on trends in electronics and informatics. IEEE; 2019, p. 788–90.
[42] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016, p. 785–94.
[43] Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning. 2010, p. 807–14.