energies

MDPI

Article

Optimal Load Distribution of CHP Based on Combined Deep Learning and Genetic Algorithm

Anping Wan [1], Qing Chang [1,2], Yinlong Zhang [3], Chao Wei [3,*], Reuben Seyram Komla Agbozo [4] and Xiaoliang Zhao [5]

1 Department of Mechanical Engineering, Zhejiang University City College, Hangzhou 310015, China
2 College of Mechanical Engineering, Anhui University of Science and Technology, Huainan 232001, China
3 Huadian Electric Power Research Institute, Hangzhou 310030, China
4 State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China
5 School of Software Technology, Zhejiang University, Ningbo 315000, China
* Correspondence: chao-wei@chder.com

Abstract: In an effort to address the load adjustment time in the thermal and electrical load distribution of thermal power plant units, we propose an optimal load distribution method based on load prediction among multiple units in thermal power plants. The proposed method utilizes optimization by attention to fine-tune a deep convolutional long-short-term memory network (CNN-LSTM-A) model for accurately predicting the heat supply load of two 30 MW extraction back pressure units. First, the inherent relationship between the heat supply load and thermal power plant unit parameters is qualitatively analyzed, and the influencing factors of the power load are screened based on a data-driven analysis. Then, a mathematical model for load distribution optimization is established by analyzing and fitting the unit's energy consumption characteristic curves on the boiler and turbine sides. Subsequently, by using a randomly chosen operating point as an example, a genetic algorithm is used to optimize the distribution of thermal and electrical loads among the units. The results showed that the combined deep learning model has a high prediction accuracy, with a mean absolute percentage error (MAPE) of less than 1.3%. By predicting heat supply load variations, the preparedness for load adjustments is done in advance. At the same time, this helps reduce the real-time load adjustment response time while enhancing the unit load's overall competitiveness. After that, the genetic algorithm optimizes the load distribution, and the overall steam consumption rate from power generation on the turbine side is reduced by 0.488 t/MWh. Consequently, the coal consumption rate of steam generation on the boiler side decreases by 0.197 kg (coal)/t (steam). These described changes can greatly increase the power plant's revenue by CNY 6.2673 million per year. The thermal power plant used in this case study is in Zhejiang Province, China.

Keywords: deep learning; load prediction; load distribution; genetic algorithm; combined heat and power

Citation: Wan, A.; Chang, Q.; Zhang, Y.; Wei, C.; Agbozo, R.S.K.; Zhao, X. Optimal Load Distribution of CHP Based on Combined Deep Learning and Genetic Algorithm. *Energies* **2022**, *15*, 7736. https://doi.org/10.3390/en15207736

Academic Editors: Qinge Xiao, Kangkang Xu and Jaroslaw Krzywanski

Received: 6 September 2022
Accepted: 12 October 2022
Published: 19 October 2022

1. Introduction

Carbon emission reduction measures have become a hot issue in China, particularly under the strategic guidance of achieving the goal of carbon peak and carbon neutrality [1]. As a result, efforts to increase renewable energy sources such as wind power and photovoltaics have surged in recent years. However, these renewable energies have their own challenges. Some of these challenges include significant randomness, drastic fluctuations, and intermittent changes with time cycles, resulting in peaks and valleys in the power grid. Moreover, traditional thermal power units present a common trend, which is the peak adjustments of renewable energy through grid scheduling—especially with increasing grid load instability [2].

In China, about half of the traditional thermal power plants are concentrated in the northern part of the country. Among these, more than 80% of them are combined heat and power (CHP) plants. CHP can be described as the crafting process of producing electricity and heat at the same time. As compared to independent heat or electricity production, co-generation improves the energy efficiency of power plants, reduces carbon emissions, and enhances the flexibility of operations [3]. Presently, the thermal power generation industry has adopted CHP as a development strategy. At the same time, it has also been positioned as the driving force behind the flexible adjustment of thermal power generation. In addition, the proper management and operation of the power plant can contribute to reducing the cost of generating electricity for the enterprise, thereby increasing the overall profit. Hence, it is significant to consider the reasonable distribution of heat and power load among the units and tap the co-generation unit's energy-saving potential [4]. Furthermore, load forecasting must be linked to the unit power generation according to the "Detailed Rules for the Implementation of Energy-Saving Power Generation Scheduling Measures". Thus, ensuring the stability of the grid load requires a high response speed of the thermal and electrical load in the peak-shaving unit. Likewise, the optimal distribution model of the thermal and electric load of the unit requires an accurate prediction of the total load demand.

The load distribution optimization method of CHP has received extensive attention in recent years. For example, Yan and Wu et al. [5] used Ebsilon software to simulate the thermal power plant system to obtain the heating domain of the unit and then used a genetic algorithm approach to optimize the distribution of thermal and electrical loads. With this method, they understood the influence of load distribution on the overall energy consumption characteristics of the thermal power plant. Second, Wang and Liu et al. [6] used a particle swarm optimization algorithm to optimize the distribution of the thermal and electrical loads among units and compared the energy consumption characteristics of power plants before and after optimization. Briefly, the above research used different optimization algorithms to improve load distribution.

Additionally, Wu and Lai et al. [7] proposed a new energy consumption calculation method—combined with economic research—and then used an adaptive genetic algorithm to solve the optimal load distribution method for the cost and profit of the CHP enterprise in real-time. Guo and Li et al. [8] developed an online performance prediction model of thermal power plants based on actual operating data and structural principles combined with unit characteristic differences and an economic analysis. They subsequently established a plant-wide operating economic benefit evaluation model and used a particle swarm optimization algorithm to obtain the multi-mode load distribution optimization scheme among heating units. Xu [9] established the functional relationship of heat rate with different main steam flow rates and different thermal power load allocations based on the actual operation data of the power plant and the theoretical model. The above research puts forward a new calculation method of unit energy consumption, and on this basis, uses an optimization algorithm to optimize the distribution of unit load.

Furthermore, Wang and Liu et al. [10] put forward a new mathematical decision-making model for load distribution based on rapidity and economy through the non-dimensional treatment of the objective function. By adjusting the size of one of the weights, the model can quickly change the speed and economic indicators of plant-level load distribution. Consequently, it minimizes coal consumption while power plants meet the essential requirement of a fast response to grid load demand.

Although the above studies have achieved good results for load optimization distribution, they have not considered the influence of the thermoelectric load adjustment response time. Zhu and Feng et al. [11] considered multi-objective indicators, including the response time of the electric load of the unit when establishing the unit load distribution model; however, when solving the multi-objective function problem through the chaotic genetic algorithm, a Pareto solution appears.

The current studies on the load distribution of CHP units in China focus on the improvement of the optimization algorithm or the method of obtaining the energy consumption curve of the unit. Unfortunately, less attention has been paid to the influence of the response time of the heat and power load adjustment [12]. In this regard, this paper considers the load adjustment time when distributing the heat and power load of the co-generation unit. In addition, this paper considers the load adjustment time when distributing the heat and power load of the co-generation unit. These are:

1. The prediction of the thermoelectric load in the next stage of the thermoelectric unit based on the combined deep learning model.
2. The construction of the energy consumption characteristic curve per unit and the optimization model of the thermoelectric load using accurate prediction.
3. Using a genetic algorithm to find the optimal solution to ensure that the load dispatching response time of the co-generation unit meets the grid dispatching requirements.

2. Deep Learning Model and Predictions

During the operation of the CHP unit, the heating load quickly changes by adjusting the steam turbine's medium and the low-pressure extraction valves. Here, the heating load's adjustment response time is much shorter than the electric load's. Hence, after receiving the grid dispatch, the electric load is regulated to correlate to the adjustment time of the power plant. The CHP follows the production principle of "power determined by heat", which means the demand for the heating load in the next stage can be predicted. This makes it easier to determine the change in the power load, prepare for the power load adjustment in advance, and reduce the response time to adjust the power load in real-time. The shorter the load adjustment time is within the allowable range of the units, the stronger the overall competitiveness of the unit load. This enables the power plant unit to have a greater load when the power grid performs load scheduling, thereby improving the economic impact of the CHP units.

In the actual production process, load demand forecasting is a prerequisite for load distribution [13]. This is the only way that production can be of benefit in a practical sense for optimal scheduling, based on which enough time can be set aside for power plant operation and maintenance planning. Hence, making reliable predictions is crucial for production deployment.

In recent years, the rapid development and wide application of deep learning have made it a hot spot in the field of load forecasting. The most common techniques include deep neural network (DNN), deep belief network (DBN), autoencoder (AE), convolutional neural network (CNN), and recurrent neural network (recurrent neural network, RNN), as well as many others; however, in the face of long-term sequence or multi-dimensional input data, a single network model still has problems with the loss of sequence feature information, the disorder of data structure information, and insufficient multi-dimensional feature mining. Thus, the combined deep learning prediction model is used to combine different deep learning models to predict the load demand for the next time period. By combining the characteristics and advantages of different models, the prediction results can be consistent with the actual situation. In general, the prediction accuracy of the combined model is higher than that of the single model [14]. Hence, to forecast the heating load demand, this paper proposes a short-term load forecasting method based on CNN-LSTM-A and XGBoost feature selection.

2.1. XGBoost Feature Selection

XGBoost can extract feature variables that have a non-linear relationship with the prediction target, and the idea of the XGBoost algorithm is to generate multiple tree models with different forms by splitting different features. Through training, new decision trees are continuously added to fit errors and improve the prediction accuracy [15]. Each feature's importance score can be directly obtained for variable feature selection after the boosting tree has been created. Subsequently, the feature parameters from the input feature set can be chosen through the importance score screening of XCBoost if they have a strong correlation with the prediction outcomes. This is followed by setting the parameter dimension of the features and the overfitting probability of the model is reduced. At the same time, the data mining capability and the prediction accuracy of the model are improved.

XGBoost based on a GBDT algorithm has very good performance. First, XGBoost explicitly adds a regular term to control the complexity of the model, which is beneficial to prevent overfitting and improve the generalization ability of the model; second, XGBoost uses second-order Taylor expansion, which can more accurately approximate the real loss function; third, XGBoost can automatically learn the processing strategy for missing values; and finally, XGBoost pre-orders each feature according to the eigenvalue and stores it as a block structure. When splitting nodes, multi-threading can be used to find each feature in parallel The best split point can greatly improve the training speed. Therefore, this paper adopts the XGBoost model with a better performance for feature extraction.

During this process, the heating load is affected by a variety of factors, thereby resulting in its strong volatility and non-linearity. Moreover, the feature parameter set has the characteristics of both multi-dimensions and multi-types. Therefore, if the input feature parameter set is not processed and directly fed into the model training, the problem of dimensional disaster occurs, resulting in the reduction of prediction accuracy and the interpretability of the model.

Thus, to verify the validity of the proposed model, a CB30-13.24/3.5/0.981-type extraction and back-pressure steam turbine unit in Zhejiang Province was used as the research subject. The daily operation data of the unit were collected from this unit. In all, a dataset of 250 dimensional units was collected from 30 June 2020, to 30 June 2021. The dimension of the data collected exceeded the appropriate dimension of the deep learning input, which is required to reduce the dimension of the data to reduce the difficulty of the learning task. As a result, a physical mechanism analysis was used to exclude the characteristic parameters that have no direct impact on the heating load. This is followed by using the XGBoost algorithm to measure the degree of correlation between each attribute parameter and the target feature.

In XGBoost, the important hyperparameters include learning_rate, max_depth, and n_estimators, where learning_rate is the learning rate, max_depth controls the depth of the tree, and n_estimators control the number of weak learners. According to experience, the value of learning_rate is set as 0.08 and the value of max_depth is set as 21. We fixed the two unchanged parameters, adjusted the value of n_estimators, and obtained nine sets of data, as shown in Table 1. Furthermore, the root mean square error (RMSE) is shown in Table 1.

Table 1. n_estimators adjusted errors and scores.

n_Estimators	RMSE	R^2 Score
150	0.0094748	0.999999423
200	0.0094693	0.999999424
250	0.0094687	0.999999424
255	0.0094686	0.999999424
260	0.0094685	0.999999424
265	0.0094685	0.999999424
270	0.0094685	0.999999424
290	0.0094685	0.999999424
300	0.0094685	0.999999424

Based on this set of data, a relationship diagram is drawn, as shown in Figure 1.

Table 1 and Figure 1 show that when the learning_rate and max_depth are fixed and the value of n_estimators is 260, the RMSE reaches the first inflection point of convergence and the R^2 score converges. Thus, the best value for n_estimators is 260.

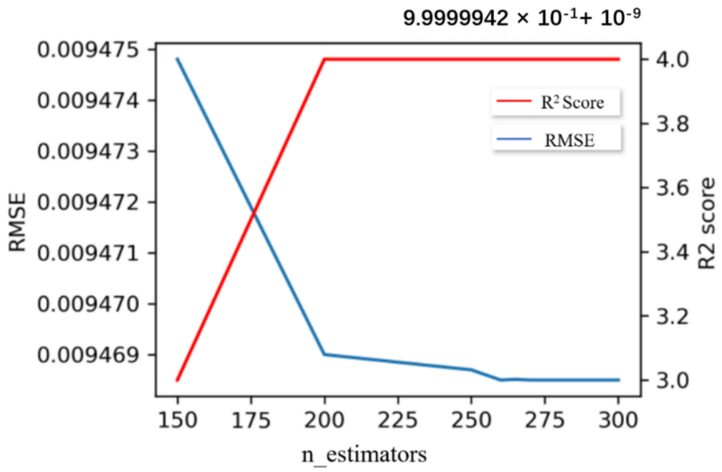

Figure 1. n_estimators adjustment diagram.

In the same way, we fixed the values of learning_rate and n_estimators and adjusted the value of max_depth to get the corresponding error and score, as shown in Table 2.

Table 2. max_depth adjustment corresponding error and score.

Max_Depth	RMSE	R^2 Score
10	0.012836066	0.999998942
15	0.010169256	0.999999336
19	0.009712356	0.999999394
20	0.009432302	0.999999429
21	0.009468482	0.999999424
23	0.009598712	0.999999408
25	0.009757911	0.999999389

A relationship diagram can be obtained from this data set, as shown in Figure 2.

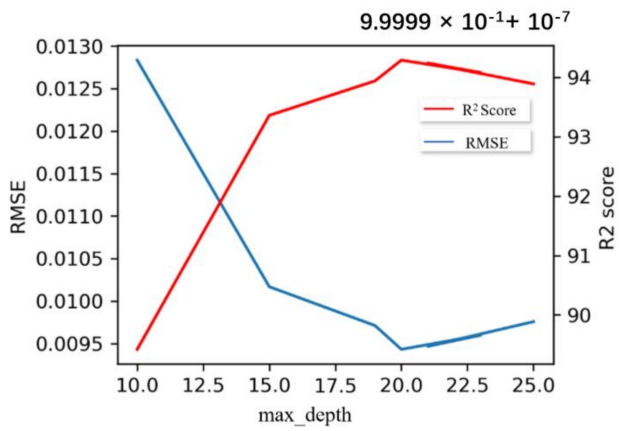

Figure 2. max_depth adjustment diagram.

Table 2 and Figure 2 show that when n_estimators and learning_rate are fixed and max_depth is 20, the RMSE and R^2 score take a minimum value. Thus, the best value for max_depth is 20.

Similarly, we fixed the values of max_depth and n_estimators, and adjusted the value of learning_rate to get the corresponding error and score, as shown in Table 3.

Table 3. learning_rate adjusts corresponding error and score.

Learning_Rate	RMSE	R^2 Score
0.02	0.100070676	0.999935703
0.03	0.008977868	0.999999482
0.04	0.006271166	0.999999747
0.05	0.007835362	0.999999606
0.06	0.008827865	0.999999587
0.07	0.00935316	0.999999438
0.08	0.009432302	0.999999429
0.09	0.010331611	0.999999315
0.1	0.010718653	0.999999262

A relationship diagram can be obtained from this set of data, as shown in Figure 3.

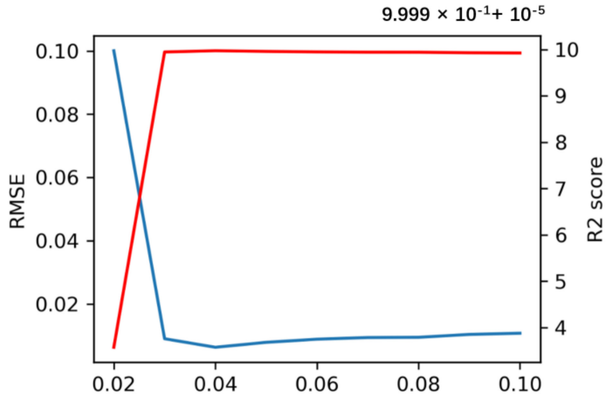

Figure 3. learning_rate adjustment diagram.

Table 3 and Figure 3 show that when n_estimators and max_depth are fixed and learning_rate is 0.04, the RMSE converges at the first inflection point and the R^2 score converges at 1. Therefore, 0.04 is the best value for learning_rate.

Following the selection of the optimal parameters, the original heating load data are subjected to feature selection, and the resulting relationship between feature importance ranking and weight value are obtained (refer to Figure 4).

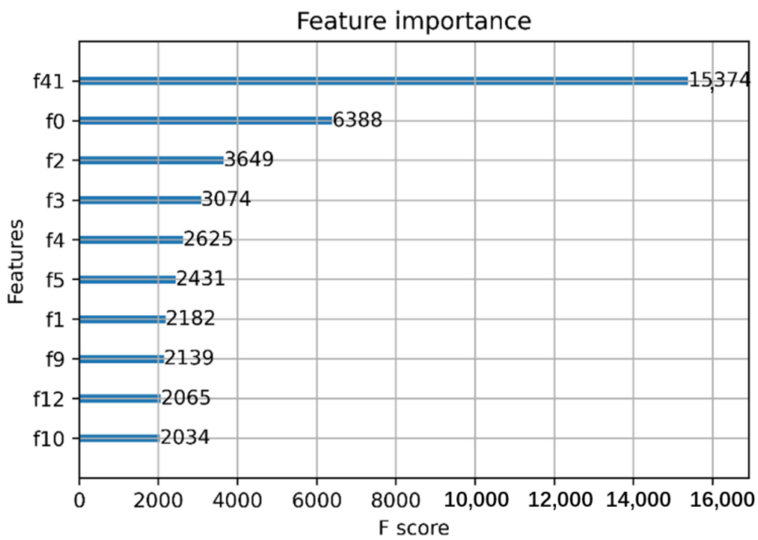

Figure 4. Feature importance ranking and weight value.

Feature importance reflects the contribution of feature parameters to the desired prediction target. The values in the figure are the influence weights of each feature. These values represent the boiler outlet steam pressure, total power generation load, total power supply load, and boiler feed water temperature—from higher to lower ranges. To avoid the curse of dimensionality, this paper selects the ten related features shown in the above figure as the input features of the hybrid deep learning model. These ten characteristics are the boiler outlet steam pressure, total power generation load, total power supply load, boiler feed water temperature, boiler feed water flow, boiler feed water pressure, power plant electricity consumption rate, instantaneous coal supply to the boiler, inlet steam flow of the main steam of the steam turbine, and the exhaust steam flow of the steam turbine.

2.2. Attention Mechanism

Since the focus of the information for each feature is different, there is no need to process each feature to predict the power load. Excessive attention to unimportant features is not conducive to obtaining important information. Therefore, a notion mechanism module was introduced to identify the important characteristics of the power load. Attention is a resource allocation mechanism that simulates the attention mechanism of the human brain, which improves the effective extraction of necessary information by cleverly and reasonably changing the attention to information, ignoring irrelevant information, and amplifying important information. In CHP power load forecasting, by calculating the probability of attention distribution, and then equally weighing it, the network selects the key information inputs for processing, which greatly improves the efficiency of the neural network. The structure of the note mechanism is shown in Figure 5.

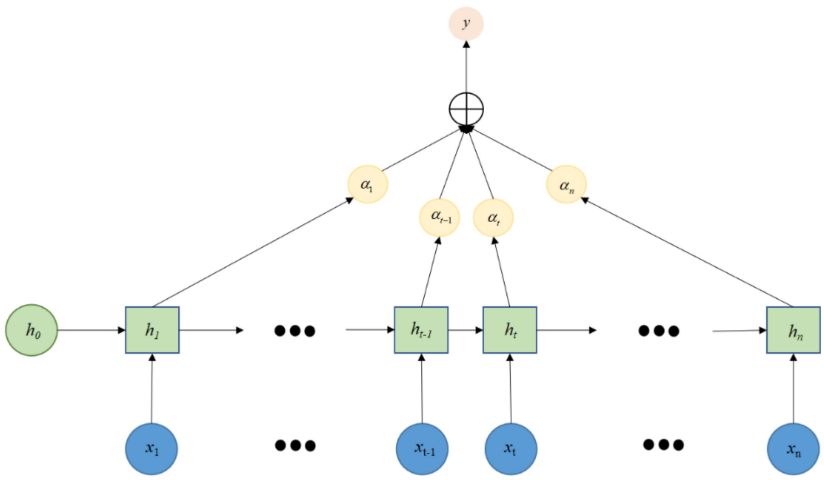

Figure 5. Attention mechanism structure.

In Figure 5, x_t denotes the input of the LSTM model, h_t represents the output of the hidden layer obtained by LSTM, α_t is the attention probability distribution of the attention mechanism output to the LSTM hidden layer, and y is the output of the LSTM with optimization by the attention mechanism.

The attention layer is the input of the layer, the output vector is processed by the LSTM activation layer, and the weight coefficient of the note layer can be calculated as:

$$e_t = utanh(wh_t + b) \tag{1}$$

$$\alpha_t = \frac{exp(e_t)}{\sum_{j=1}^{t} e_j} \tag{2}$$

$$s_t = \sum_{t=1}^{i} \alpha_t h_t \tag{3}$$

where e_t denotes the attention scoring function used to calculate the correlation between the power load and the LSTM network layer output vector h_t at moment t, u and w are weight coefficients, b is the bias coefficient, and s_t represents the output of the attention layer at time t.

2.3. Data Preprocessing

The experimental data were obtained from 30 June 2020 to 30 June 2021, from the CHP plant in Zhejiang Province. The data horizon in this case was 1 min. As such, this article predicts the change in heat load over a period of one hour, making the original data extraction a total of 60 data points.

The feature data are not of the same order of magnitude, which led to an increase in the number of iterations used for finding the optimal solution of the objective function. To solve this problem and speed up the process of searching for the optimal solution by the stochastic gradient descent algorithm, we normalized the data, mapped the data features to the [0, 1] interval, and removed the dominant features of the macrodimension. Equation (1) was used to normalize the input data as shown below:

$$x_n = \frac{x_o - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

where x_n represents the normalized data, x_o represents the original data, $x_{\max}$ represents the maximum data, and $x_{\min}$ represents the minimum data.

2.4. Hybrid Deep Learning Predictive Models

In order to improve the load prediction accuracy, this paper proposes a combined load prediction model of CNN-LSTM based on the attention mechanism. This method uses a CNN to extract effective feature vectors from historical load sequences and uses an LSTM network to model and learn the dynamic changes of the features proposed by the CNN. It also introduces the attention mechanism to give LSTM hidden states different probability weights to strengthen the impact of important information on the load demand. The brief principle of combining deep learning models is as follows:

A convolutional neural network (CNN) uses local connections and weight-sharing to perform higher-level and more abstract processing on the original data, which can effectively and automatically extract the internal features in the data [16]. The long-short-term memory network (LSTM) on the other hand is an improved version of the recurrent neural network (RNN) with a gradient explosion or gradient disappearance being used during training. Research and applications show that LSTM is more effective than other deep learning algorithms when dealing with time series problems [17]. The attention mechanism is a mechanism that imitates the way the human brain uses and allocates attention information resources. The core idea is to subtly change people's perception of attention to information, ignore redundant and useless information, and amplify the required information. This dramatically improves the design accuracy of the model [18]. The attention mechanism effectively improves the problem that LSTM misses with sequence information because the input time series is too long. The method of randomly assigning probability is adopted instead of the original method of randomly assigning weights.

Here, the historical heating data are used as inputs to extract features through the CNN layer. The LSTM layer and the attention layer learn the internal variation law of heating loads from the extracted features to achieve an accurate prediction. Subsequently, the output of the LSTM layer and the weights operated by the attention layer are transformed into a one-dimensional structure by the fully connected layer. After that, the predicted sequence releases the results through the output layer. The proposed research framework for CHP heating load forecasting is shown in Figure 6, which includes the input layer, CNN layer, LSTM layer, attention layer, and output layer.

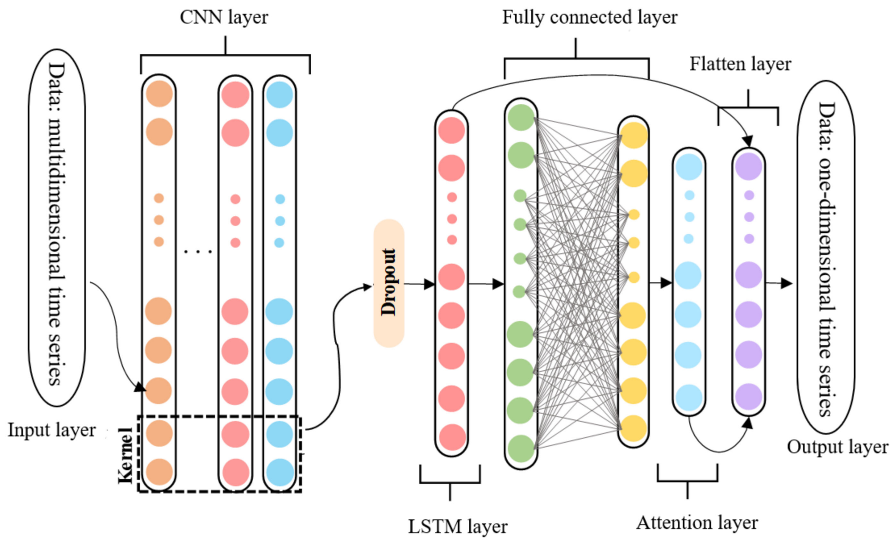

Figure 6. Structure diagram of hybrid deep learning model.

Figure 7 shows the change process of MSE with different mini-batches when the number of epochs is 30 and the data are sent to the model for training after normalization.

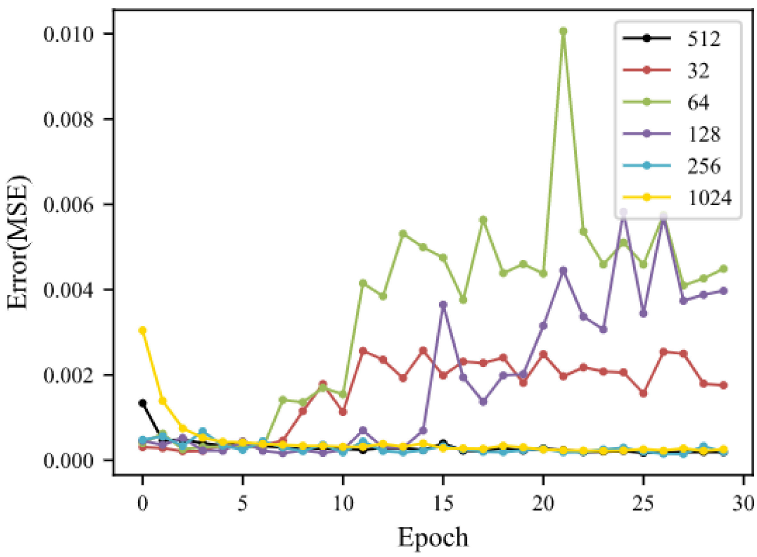

Figure 7. Model MSE change.

It can be seen that when the value of mini-batches is 256, the convergence speed of the model is the fastest and the error is the smallest.

Python was used to implement the CNN-LSTM model with an attention mechanism optimization forecasting model, which was done within the TensorFlow 2.0 framework. The experimental hardware was configured as NVIDIA(R) GeForce(R) MX250, GPUIntel (R) Core (TM) i7-10510U CPU with 16G memory. The network parameters were updated using the Adam optimization algorithm, and the initial learning rate was set to 0.01. The set neural network training parameters are as follows: the maximum number of iteration epochs is 30 and the mini-batch size is 256 [19]. The sample set is divided into training samples and test samples at a ratio of 4:1, and the specific structural parameters of the model are shown in Table 4.

Table 4. The structure parameters of the model.

Layer Category	Neurons	Remark
Input layer	24 × 11	
Convolution layer	24 × 64	Kernel 1 × 11/1
Dropout layer	24 × 64	Retention 0.8
LSTM layer	24 × 128	
Fully connected layer	24 × 128	
Attention layer	24 × 128	Softmax
Flatten layer	1 × 5120	
Output layer	1 × 24	

The predicted time interval is 1 h. In order to accurately mine the characteristics of the data within a day cycle, we use 24 data points as time steps to predict the target data. The original time series data has a total of 9 dimensions (24 × 9), which are input through the input layer, and the data features were extracted by the convolutional layer by down-sampling the max pooling layer. In the process, the dropout layer discards 20% of the neurons to prevent overfitting. Finally, the feature sequence is converted into a 24 × 64 vector and input into the LSTM layer. To calculate the attention distribution, the output of the LSTM layer is merged into the attention layer through the fully connected layer. Next, the weight distribution optimization is carried out. Furthermore, there are a set of output prediction results for each 24-time step.

3. Units Load Characteristics of CHP

The load characteristic curves of the boiler side and the turbine side were generated by fitting and an appropriate mathematical model was established based on the difference in the load characteristics after examining the actual operation data of the co-generation unit. Thereafter, the genetic algorithm optimizes the objective function for the actual thermal and electrical load demand of the CHP plant. This optimization process helps in achieving the optimal steam consumption rate of power generation on the turbine side and the optimal coal consumption rate of steam production on the boiler side. Consequently, energy is saved and consumption is reduced.

3.1. Full Load Characteristics Curve of Boiler

The required thermal parameters and energy consumption index parameters are obtained through the performance test of each boiler, and finally, the load characteristic equation of each boiler is obtained by the method of curve fitting. The system has three UG-220/13.7-Ms (boiler model), which produce 220 t/h of high temperature and high pressure. The test points were arranged, and boiler performance tests were conducted on the #1, #2, and #3 furnaces under seven load states, ranging from 110 to 240 t/h, to obtain the boiler's total combustion performance data. The test content includes coal sampling and coal quality analysis, fly ash sampling, flue gas composition measurement, bottom slag sampling, and temperature measurements. The inverse balance method was used to determine the boiler's efficiency based on test-related parameters, including the exhaust gas temperature, ash carbon content, coal quality, and composition of the exhaust gases. To illustrate this clearly, the boiler efficiencies of the three boilers under seven load states between 110 and 240 t/h are shown in Table 5.

Furthermore, polynomial fitting was performed on the boiler efficiencies with the origin software, with the fitting curve equation given below in Equation (5):

$$\eta = \text{Intercept} + \mathrm{B}_1 \cdot D_b + B_2 \cdot D_b^2 \tag{5}$$

In Equation (5), η is the boiler efficiency, D_b is the boiler steam production, and Intercept, B_1, and B_2 are the polynomial coefficients; the values are shown in Table 6. The fitting degrees were 0.59, 0.88, and 0.77, respectively.

Table 5. Boiler efficiency under load condition of three boilers (%).

Load	#1 Boiler	#2 Boiler	#3 Boiler
240 t/h	94.01	93.87	93.77
220 t/h	93.64	93.71	93.68
200 t/h	93.76	93.91	94.26
180 t/h	93.30	93.73	93.79
160 t/h	93.69	93.51	93.04
140 t/h	93.53	93.44	93.13
110 t/h	92.48	91.95	92.32

Table 6. Boiler load characteristic equation polynomial coefficient.

Coefficient	#1 Boiler	#2 Boiler	#3 Boiler
Intercept	89.506	85.433	87.078
B_1	0.038	0.083	0.063
B_2	-8.40411×10^{-5}	-2.02479×10^{-4}	-1.47399×10^{-4}

3.2. Full Load Characteristics Curve of Steam Turbine

To calculate the steam turbine's grid performance under various steam extraction–exhaust combined operating settings, the historical operation and maintenance data from the CHP plant is fitted into full load characteristics curves of the steam turbine. This leads to a reduction in the manual workload, where the cost and more reliable data are provided for the subsequent optimal allocation of loads. As such, the medium-pressure extraction steam flow G_g and the low-pressure exhaust steam flow G_p are calculated daily, along with the turbine inlet steam flow G and power generation N.

During the fitting process of the characteristic equation, in addition to the influence of the medium-pressure extraction steam and the low-pressure exhaust steam on the equation fitting, it is also related to whether the high-temperature heater is operational. The characteristic equation only pertains to whether the high-temperature heater is used under actual operating conditions when there is no medium-pressure extraction steam. Additionally, the system can be used to determine whether the two high-temperature heaters are in operation by using 40 °C as the dividing line for the temperature difference between the inlet and outlet of the two high-temperature heaters in each turbine. If the temperature difference is greater than 40 °C, the high-temperature heater is in operation; if it is less than 40 °C, the high-temperature heater is not in operation. Through fitting and analysis, the characteristic equations between the steam inlet volume, power generation power, medium-pressure extraction steam, and low-pressure exhaust steam of the steam turbine are calculated, as represented in the equations below:

$$G = f(G_p, G_g) \tag{6}$$

$$N = g(G_p, G_g) \tag{7}$$

In Equations (6) and (7), G is the intake flow of the steam turbine, t/h; G_p is the steam turbine exhaust flow, t/h; $f(G_p, G_g)$ is the steam turbine extraction flow; $g(G_p, G_g)$ is the steam turbine inlet steam flow characteristic equation; $g(G_p, G_g)$ is the steam turbine power generation characteristic equation; and N denotes the turbine power generation, MW.

The detailed expression of its characteristic equation is shown in Table A1 [20].

4. Mathematical Model of Unit Load Optimal Distribution

4.1. Mathematical Model of Turbine Side Load Distribution

There are two heating steams with different pressures on the turbine side: a small amount of medium-pressure extraction heating steam and a low-pressure exhaust heating steam. The #1 steam turbine is primarily in charge of producing medium-pressure extraction heating steam due to its low demand. The steam intake and power output of each turbine can be changed by altering its medium-pressure extraction and low-pressure exhaust. The distribution option that minimizes the overall power generation steam consumption rate on the turbine side is designed to obtain the ideal distribution of the thermoelectric load across units in all combinations where the total amount of low-pressure exhaust steam fulfills the external heating demand.

The objective function for the minimum distribution scheme of the steam consumption rate for power generation on the turbine side is:

$$\min d = \frac{\sum_{i=1}^{m} G_i}{\sum_{i=1}^{m} P_i} = \frac{\sum_{i=1}^{m} f(G_{pi}, G_{gi})}{\sum_{i=1}^{m} g(G_{pi}, G_{gi})} \tag{8}$$

where d denotes the steam consumption rate of power generation; m denotes the turbine quantity; G_i denotes the steam intake of the i-th unit, t/h; G_{pi} denotes the low-pressure exhaust volume of the i-th unit, t/h; P_i denotes the generating power of the i-th unit; MW; and G_{pi} denotes the medium-pressure extraction of the i-th unit, t/h.

The unit capacity constraints are:

$$G_i^{\min} \leq G_i \leq G_i^{\max} \tag{9}$$

$$P_i^{\min} \leq P_i \leq P_i^{\max} \tag{10}$$

where $G_i^{\max}$ denotes the maximum allowable steam intake flow of the i-th unit, t/h; $G_i^{\min}$ denotes the allowable minimum intake steam flow of the i-th unit; $P_i^{\min}$ denotes the minimum generating power of the i-th unit, MW; and $P_i^{\max}$ denotes the maximum generating power of the i-th unit, MW.

The exhaust flow constraint is:

$$G_{gi} \leq G_{gi}^{\max} \tag{11}$$

where $G_{gi}^{\max}$ represents the maximum allowable medium-pressure heating extraction steam flow of the i-th unit, t/h.

The exhaust flow constraint is:

$$G_{pi}^{\min} \leq G_{pi} \leq G_{pi}^{\max} \tag{12}$$

where $G_{pi}^{\max}$ is the maximum allowable low-pressure exhaust steam flow of the i-th unit, t/h; and $G_{pi}^{\min}$ denotes the minimum allowable low-pressure exhaust steam flow of the i-th unit, t/h.

The heating load equation constraint is:

$$\sum_{i=1}^{m} G_{pi} = G_r \tag{13}$$

where G_r is the total heating flow, t/h.

4.2. Mathematical Model of Boiler Side Load Distribution

Once the load on the steam engine side has been distributed optimally, the steam output that corresponds to the steam engine side's steam intake volume is distributed to each boiler. This method is known as the optimal load distribution. By using the optimal load distribution approach, the boiler side's overall rate of coal consumption during the creation of steam is reduced.

The characteristic equation between the boiler efficiency and steam production has been obtained from boiler performance tests:

$$\eta = h(G_b) \tag{14}$$

where η is the boiler efficiency; $h(G_b)$ denotes the boiler coal consumption characteristic curve; and G_b denotes the boiler steam production.

The calculation formula of coal consumption is:

$$B_i = \frac{G_{bi}(H_{si} - H_{gi})}{\eta_i Q} = \frac{G_{bi}(H_{si} - H_{gi})}{h(G_{bi})Q} \tag{15}$$

where B_i is the coal consumption of the i-th boiler, t/h; G_{bi} is the steam production of the i-th boiler, t/h; H_{si} is the enthalpy of the main steam of the i-th boiler, kJ/kg; H_{gi} is the feed water enthalpy value of the i-th boiler, kJ/kg; η_i is the efficiency of the i-th boiler; and Q is the calorific value of coal, kJ/kg. The standard coal low-calorific value of 29,307.6 kJ/kg is used in the calculation.

The objective function is the lowest coal consumption rate of steam generation:

$$\min b = \frac{\sum_{i=1}^{n} B_i}{\sum_{i=1}^{n} G_{bi}} = \frac{\sum_{i=1}^{n} \frac{G_{bi}(H_{si}-H_{gi})}{\eta_i Q}}{\sum_{i=1}^{n} G_{bi}} = \frac{\sum_{i=1}^{n} \frac{G_{bi}(H_{si}-H_{gi})}{h(G_{bi})Q}}{\sum_{i=1}^{n} G_{bi}} \tag{16}$$

In Equation (16), b is the coal consumption rate for steam production and n is the number of boilers.

The boiler capacity constraints are:

$$G_{bi}^{\min} \le G_{bi} \le G_{bi}^{\max} \tag{17}$$

where $G_{bi}^{\min}$ is the minimum steam production flow of the i-th boiler, t/h; and $G_{bi}^{\max}$ is the maximum steam production flow of the i-th boiler, t/h.

5. Genetic Algorithm for Optimizing Load Distribution

A genetic algorithm has strong robustness, parallelism, and global optimization for solving multi-objective and non-linear optimization problems. It also has strong adaptability when dealing with unit load distribution problems. A genetic algorithm has been widely used in the optimization of power systems [21].

The procedures of a genetic algorithm to optimize the allocation of the thermoelectric load are as follows:

(1) Coding Binary coding is a common coding method used for genetic algorithms, but it is hard to reflect the structural characteristics of the problem using this method. For some optimization problems of continuous functions, its local search ability is poor because of the random characteristics of genetic operation. Moreover, the number of decision variables increases with the scale of the system, and subsequently, the code string will be long. This leads to the Hamming cliff problem in mutation operations [22], which increases the difficulty of finding the optimal solution. To solve this problem, this paper uses Gray code to encode individuals.

(2) Fitness calculation Since the objective function is to find the minimum value, the fitness calculation function is set as the negative of the objective function.

(3) Selection operation When the fitness value is negative, random league selection is adopted. In league selection, the number of individuals whose fitness is compared at each time is called the league size. The value of league size N in this paper is 3.

(4) Crossover operation To extend the search space while maintaining a good information exchange, this paper adopts the uniform crossover operation.

(5) Mutation operation To improve the local search ability of the genetic algorithm, maintain the diversity of the population, and prevent the premature convergence, the basic bit mutation operation is adopted.

(6) Parameter settings In the genetic algorithm, the population size M is 100, and one individual represents one chromosome. Two units on the steam turbine side represent a two-dimensional vector, and three boilers on the boiler side represent a three-dimensional vector. The maximum number of iterations $T_{\max}$ is set to 100. This is because the number of units in the load optimization allocation is small and convergence is achieved within 100 iterations. The crossover rate p_c is 1, the mutation rate p_m is 0.01, and the discrete precision is 0.01.

6. Case Study

The procedures for load distribution based on predictions are shown in Figure 8.

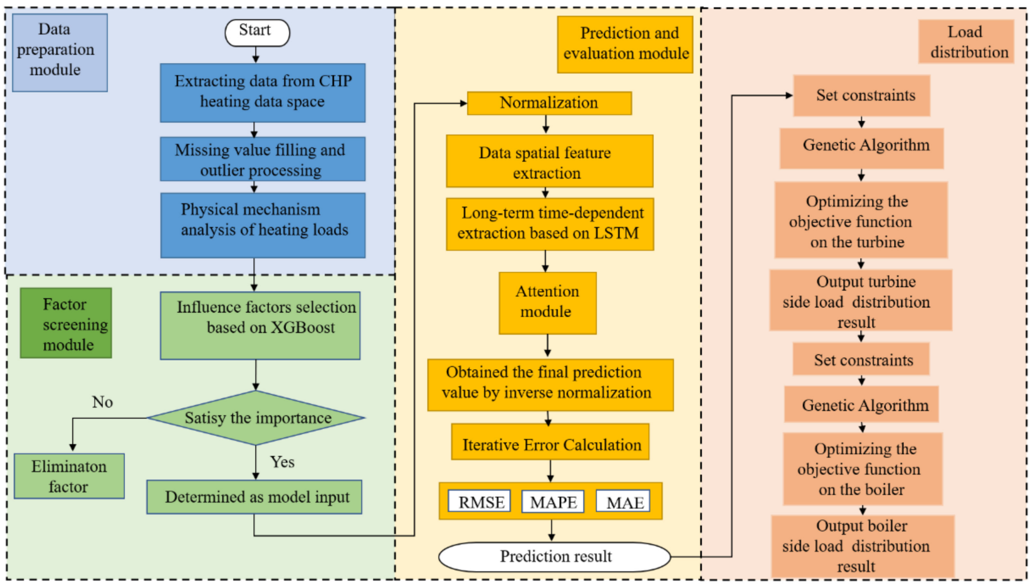

Figure 8. Research flow chart of this paper.

(1) Data preparation. The data are extracted from the data space and preprocessed; then the correlation between the power loads and influencing factors is analyzed based on a physical mechanism analysis.

(2) Influencing factor screening module. Influence factor selection is based on XGBoost.

(3) Forecasting and evaluation module. This forecasting module includes the CNN layer, LSTM layer, and attention layer.

(4) Load distribution. The prediction results are input as the constraints on the engine side, and then the objective function is optimized by the genetic algorithm, and finally, the load distribution result is produced.

In this paper, the 12-day load prediction started on 4 June 2021 and is compared with the measured data. The prediction results and their errors are shown in Figure 9. This shows that the predicted results are highly coincidental with the measured data. The RMSE of Unit 1 was 2.298 and the mean absolute percentage error (MAPE) was 1.2%. The RMSE of Unit 2 was 2.337 and the MAPE was 1.3%. The proposed CNN-LSTM-A model has a high prediction accuracy.

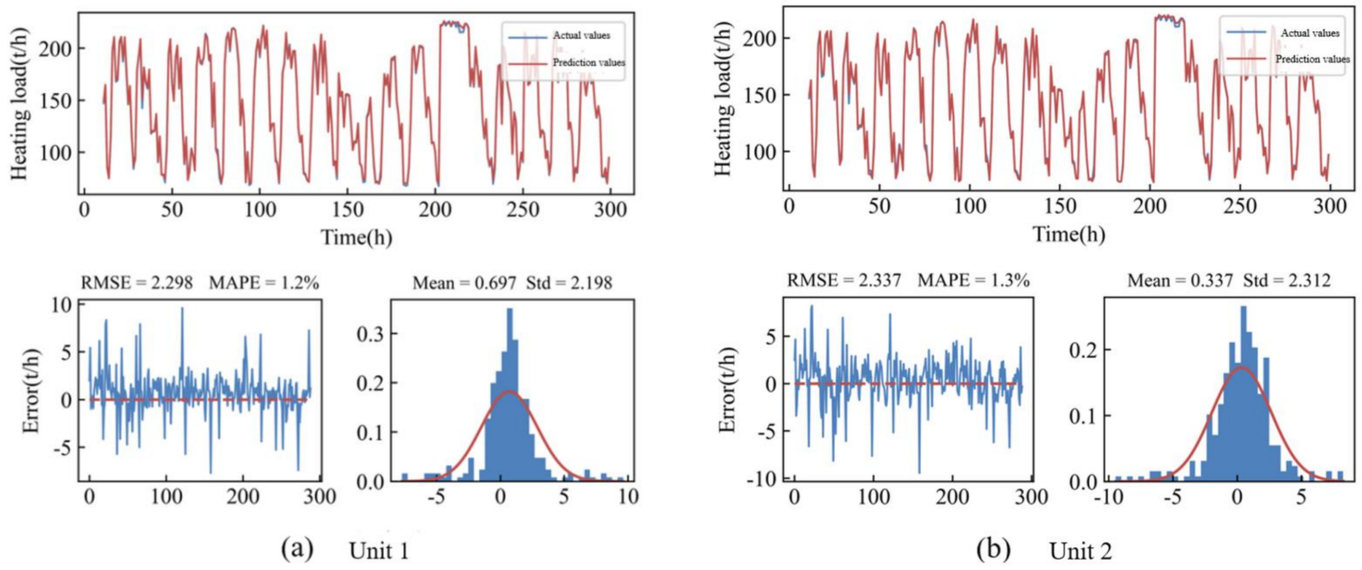

Figure 9. Forecasting and error of steam turbine: (**a**) Steam Turbine Unit No. 1; (**b**) Steam Turbine Unit No. 2.

Each model of each steam turbine unit was verified ten times, and the average prediction error is shown in Figure 10.

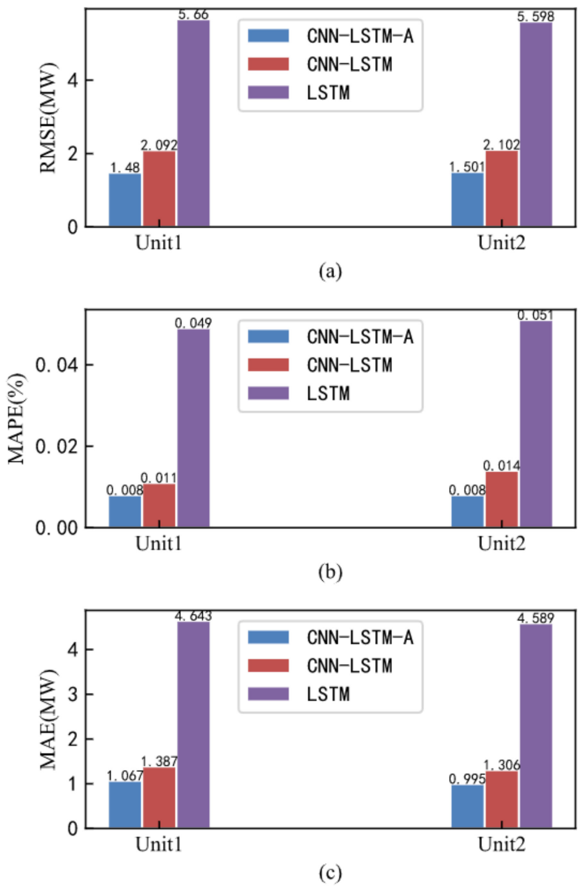

Figure 10. Prediction error: (**a**) RMSE; (**b**) MAPE; (**c**) MAE.

The CNN-LSTM-A load prediction model proposed in this paper has a minimum RMSE and a MAPE of less than 1.1% when predicting the power of the two turbines, which has a very high accuracy.

Table A2 compares the operating parameters and energy consumption indicators before and after the unit load is optimized by the genetic algorithm under the randomly selected operating points. Following the optimal load distribution, the low-pressure exhaust steam flow of the #1 turbine decreased by 14.2 t/h, while the exhaust steam flow of the #2 turbine increased by 14.19 t/h; however, the total amount of low-pressure heating flow remained constant and still met the external heating demand. Furthermore, due to the lower demand, the #1 turbine is responsible for carrying all of the medium-pressure extraction steam flow. With the variation in the thermal load, the generating power of the two steam turbines was also adjusted. As a result, the overall generating power of the unit increased by 5.51 MW while the generating power of the #2 steam turbine increased by 7.51 MW and the generating power of the #1 steam turbine decreased by 1.99 MW.

In the turbine, the steam inlet flow changes as the heat load changes. This change causes the steam inlet volume of the #1 turbine to decrease by 18.95 t/h, and the #2 turbine to increase by 44.43 t/h. Although the overall steam consumption of the unit increased by 25.48 t/h, its overall steam consumption rate decreased by 0.488 t/MWh, that is, for every 1 MWh of electricity generated, the steam consumption was reduced by 488 kg. Given this, the total steam consumption after optimal distribution on the steam turbine side is optimally distributed among the boilers. After the optimal distribution of the boiler side load, the steam production of the #1 boiler decreased by 35.71 t/h, the steam production of the #2 boiler increased by 47.91 t/h, and the steam production of the #3 boiler increased by 11.53 t/h. Hence, the overall steam production on the boiler side increased by 23.73 t/h, and the total steam production was equal to the optimized total steam intake on the turbine side. In addition, the coal consumption on the boiler side increased by 2.20 t/h, but the coal consumption rate for steam production decreased by 0.197 kg coal/t steam, that is, 197 g of coal consumption can be saved for every 1 t of steam produced.

Based on an annual operation of 6000 h, the unit price of standard coal entering the plant in the second half of 2021 is estimated to be 0.5058 CNY/kWh. The power plant can reduce CO_2 emissions by 2860.44 tons per year under the condition that it produces the same amount of steam after an optimal load distribution. Furthermore, in the case of meeting the same external heating load demand, the revenue from power generation and the cost of coal consumption will increase by CNY 16.7217 million and CNY 10.4545 million, respectively. In all, the power plant's revenue will soar by CNY 6.2673 million annually.

7. Conclusions

This paper proposed a CNN-LSTM-AM combined deep learning prediction model based on XGBoost feature selection. It indicated that the heating load of CHP units can be precisely predicted using this method. By using the full load characteristic curves of the boiler side, a mathematical model of the load optimization distribution was developed through the fitting of test data on the turbine side. Furthermore, the operating economy of the optimized unit was significantly increased by comparing the operating parameters and energy consumption indicators before and after the optimization of randomly chosen operating points. The conclusions from the experimental analysis are as follows.

(1) The proposed CNN-LSTM-AM model combines the advantages of the three single models while weakening their disadvantages. As a result, it has a low prediction error value. The average absolute percentage error is less than 1.3 percent, and it accurately predicts the unit's future heating load. The variation in the electrical load can be identified based on the precise heating load prediction value of the combined model, with the thermal load adjustment planned in advance. In addition, it decreases the response time for real-time electrical load adjustment and leaves enough time for the power plant's operation and maintenance. As a result, when the power grid schedules loads, the unit will receive a larger load, thereby enhancing the power plant's ability to compete in the market.

(2) The load adjustment response problem becomes a load forecasting problem when the load adjustment response time is taken into account. The load distribution is split into the turbine side and the boiler side at the same time. This reduces the complexity of

optimizing the objective function, avoids solving the multi-objective problem, and increases the effectiveness of solving.

(3) On the turbine side, the overall steam consumption rate of energy production was decreased by 0.488 t/MWh, and on the boiler side, the overall coal consumption rate of steam production was decreased by 0.197 kg coal/t steam. This is a result of the comparisons of the energy consumption indices between the before and after of the load distribution optimization and the randomly selected operating points. The power plant can then produce the same amount of steam while improving the energy efficiency and reducing CO_2 emissions by 2860.44 tons annually. The thermal power plant can also increase revenue by CNY 6.2673 million annually, thereby improving the economic viability of the unit.

Author Contributions: Conceptualization, Q.C. and A.W.; methodology, Q.C. and A.W.; software, Q.C.; validation, Q.C. and C.W.; formal analysis, Q.C. and Y.Z.; investigation, R.S.K.A. and C.W.; resources, A.W. and C.W.; data curation, A.W. and Y.Z.; writing—original draft preparation, Q.C.; writing—review and editing, R.S.K.A.; visualization, Q.C. and X.Z.; supervision, C.W. and A.W.; project administration, Y.Z. and X.Z.; funding acquisition, A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (51705455), Special support for marine economic development of Guangdong Province(GDNRC[2022]28).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the useful comments from editors and reviewers.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Piecewise fitting load characteristic equation of turbine unit.

Steam Turbine	Project	Medium Pressure Extraction	High Temperature Heater	Piecewise Fitting Functions	MAPE
#1 Steam turbine	Steam intake	With medium-pressure extraction $Gg > 0$ t/h	/	$G = 76.76539 + 0.1681\,G_g + 0.31973\,G_p + 4.69726 \times 10^{-4}\,G_g^2 + 0.00246\,G_p^2 + 0.00479\,G_gG_p$	1.72%
		No medium-pressure extraction $Gg = 0$ t/h	High temperature heater not working	$G = 1.0013\,G_p - 0.49921$	
			High temperature heater working	$G = 1.24028\,G_p - 5.71241$	
	Power generation	With medium-pressure extraction $Gg > 0$ t/h	/	$N = -14.15576 + 0.01318\,Gg + 0.2177\,Gp + 1.38881 \times 10^{-4}\,Gg^2 - 2.26418 \times 10^{-5}\,G_p^2 + 5.71866 * 10^{-4}\,G_gG_p$	2.38%
		No medium-pressure extraction $Gg = 0$ t/h	High temperature heater not working	$N = 0.1604\,G_p - 7.8768$	
			High temperature heater working	$N = 0.18626\,G_p - 8.62791$	

Table A1. *Cont.*

Steam Turbine	Project	Medium Pressure Extraction	High Temperature Heater	Piecewise Fitting Functions	MAPE
#2 Steam turbine	Steam intake	With medium-pressure extraction $Gg > 0$ t/h	/	$G = 37.94147 + 1.7567\ G_g + 0.47466\ G_p - 0.00294\ G_g^2 + 0.00253\ G_p^2 - 0.00195\ G_gG_p$	1.53%
		No medium-pressure extraction $Gg = 0$ t/h	High temperature heater not working	$G = 0.79017\ G_p + 30.29029$	
			High temperature heater working	$G = 1.23827\ G_p - 5.55379$	
	Power generation	With medium-pressure extraction $Gg > 0$ t/h	/	$N = -10.03398 + 0.26538\ G_g + 0.14488\ G_p - 4.98094 \times 10^{-4}\ G_g^2 + 2.1796 \times 10^{-4}\ G_p^2 - 4.98446 \times 10^{-4}\ G_gG_p$	2.49%
		No medium-pressure extraction $Gg = 0$ t/h	High temperature heater not working	$N = 0.12973\ G_p - 3.71356$	
			High temperature heater working	$N = 0.18277\ G_p - 8.00891$	

Table A2. Genetic algorithm-based load optimization distribution results.

Parameter	Unit	Before Optimization	After Optimization	Difference
		Turbine side		
#1 turbine main steam inlet flow	t/h	141.35	122.40	−18.95
#2 turbine main steam inlet flow	t/h	303.23	347.66	44.43
#1 turbine power generation	MW	12.87	10.88	−1.99
#2 turbine power generation	MW	34.8	42.31	7.51
#1 turbine medium pressure heating extraction steam flow	t/h	61.51	61.51	0
#2 turbine medium pressure heating extraction steam flow	t/h	0	0	0
#1 turbine exhaust flow	t/h	117.53	103.33	−14.2
#2 turbine exhaust flow	t/h	208.48	222.67	14.19
Power generation	MW	47.67	53.18	5.51
Steam consumption	t/h	444.58	470.06	25.48
Low-pressure steam	t/h	326	326	0
Medium pressure steam	t/h	61.51	61.51	0
Power generation steam consumption rate	t/MWh	9.326	8.838	−0.488
		Boiler side		
#1 Boiler outlet steam flow	t/h	145.9	110.19	−35.71
#2 Boiler outlet steam flow	t/h	150.42	198.33	47.91
#3 Boiler outlet steam flow	t/h	150.01	161.54	11.53
Steam production	t/h	446.33	470.06	23.73
Coal consumption	t/h	39.74	41.76	2.02
Coal consumption rate in steam prduction	kg (coal)/t (steam)	89.037	88.840	−0.197

References

1. He, M.S. Feasibility Study on Operation for Power Peaking of Cogeneration Units. Master's Thesis, North China Electric Power University, Beijing, China, 2012.
2. Zhang, Y.B. Study on operation optimization of thermoelectric units based on multiple heating modes. *Shanxi Electr. Power* **2020**, *5*, 52–57.
3. Wan, Y.; Sun, S.M.; Ge, Z.H.; He, J.R. Thermal-Economic Analysis of High Back Pressure Heating Retrofit for Large-Scale Cogeneration Unit under Full Condition. *Electr. Power Constr.* **2016**, *37*, 131–137. [CrossRef]
4. Jia, S.; Kang, X. Multi-Objective Optimal Scheduling of CHP Microgrid Considering Conditional Value-at-Risk. *Energies* **2022**, *15*, 3394. [CrossRef]

5. Yan, X.S.; Wu, Z.H.; Yin, G.; Wu, R.J.; Guo, L.D. Research on optimization of heating region and load distribution mode of heating units. *Chin. Meas. Test Technol.* **2022**, *48*, 148–153.
6. Wang, S.; Liu, M.; Yan, J.J. Optimizing heat-power load distribution of thermal power plants based on particle swarm algorithm. *J. Xi'an Jiaotong Univ.* **2019**, *53*, 159–166.
7. Wu, T.; Lai, F.; Liu, Z.; Wang, Z.W.; He, X.; Xu, C.G.; Xue, H.G.; Shi, L.; Xu, C. Intelligent load distribution of cogeneration units in deep peak regulation mode. *Therm. Power Gener.* **2021**, *50*, 119–127. [CrossRef]
8. Guo, L.J.; Li, B.; Chen, L.D. Study on Load Distribution Optimization and Operation Economy Improvement of Multiple Heating Units in Thermal Power Plant. *J. Eng. Therm. Energy Power* **2022**, *37*, 18–24. [CrossRef]
9. Xu, Z.S. Optimization Research of Electric Load and Heating Load between Units Power Supply Based on Genetic Algorithm. Master's Thesis, Jilin University, Changchun, China, 2016.
10. Wang, Z.G.; Liu, J.Z.; Tan, W.; Yang, G.J. Multi-objective Optimal Load Distribution Based on Speediness and Economy in Power Plants. In Proceedings of the CSEE, Guiyang, China, 25–27 May 2013; Volume 19, pp. 86–92.
11. Zhu, Y.; Li, Q.J.; Feng, Y.X. Load optimal distribution for double extraction cogeneration unit based on comprehensive economic cost index. *Therm. Power Gener.* **2014**, *43*, 15–20+26.
12. Li, X.B. Study of Multi-objective Optimization and Multi-attribute Decision Making of Economic Load Dispatch Problem. *Proc. CSEE* **2008**, *28*, 102–107.
13. Zhou, Z.; Jiao, W.L.; Ren, L.M.; Tian, X.H. Combined forecasting model of gas daily load based on weight distribution of ant colony algorithm. *J. Harbin Inst. Technol.* **2021**, *53*, 177–183.
14. Zhu, J.L.; Xun, Z.H.; Wang, Y.X.; Cui, Q.; Chen, W.Y.; Lou, J.C. Short-term Power Load Forecasting Based on CNN-BiLSTM. *Power Syst. Technol.* **2021**, *45*, 4532–4539. [CrossRef]
15. Hu, M.; Hu, X.; Deng, Z.; Tu, B. Fault Diagnosis of Tennessee Eastman Process with XGB-AVSSA-KELM Algorithm. *Energies* **2022**, *15*, 3198. [CrossRef]
16. Zou, Y.; Feng, W.; Zhang, J.; Li, J. Forecasting of Short-Term Load Using the MFF-SAM-GCN Model. *Energies* **2022**, *15*, 3140. [CrossRef]
17. Osowski, S.; Szmurlo, R.; Siwek, K.; Ciechulski, T. Neural Approaches to Short-Time Load Forecasting in Power Systems—A Comparative Study. *Energies* **2022**, *15*, 3265. [CrossRef]
18. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
19. Rao, M.; Wang, L.; Chen, C.; Xiong, K.; Li, M.; Chen, Z.; Dong, J.; Xu, J.; Li, X. Data-Driven State Prediction and Analysis of SOFC System Based on Deep Learning Method. *Energies* **2022**, *15*, 3099. [CrossRef]
20. Xu, K. Thermal System Modeling of Main-Pipeline Cogeneration Unit and Combined Heat and Power Optimized Distribution. Master's Thesis, Zhejiang University, Hangzhou, China, 2020. [CrossRef]
21. Liu, L.H.; Han, P.; Wang, D.F. A new chaos genetic algorithm and its application in function optimization. *J. North China Electr. Power Univ.* **2010**, *37*, 93–96.
22. Tan, T.L.; Zhang, R. Reactive power optimization based on genetic/tabu research hybrid algorithm. *Power Syst. Technol.* **2004**, *11*, 57–61. [CrossRef]