Proceedings of the 4th International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.33645

Regression Prediction of Industrial Steam Boiler Thermal Efficiency Based on Machine Learning Algorithms

Jiaxi Li*, Jinpeng Han

College of Engineering, China University of Petroleum - Karamay Campus, Karamay, China
**Corresponding Author. Email: 1034745160@qq.com*

Abstract. As the core equipment of energy power system, industrial steam boiler is difficult to achieve accurate thermal efficiency prediction by traditional mechanism modeling and counter-equilibrium calculation under deep peak shaving and complex working conditions. Single machine learning and deep learning models have problems such as insufficient timing capture, parameter dependence on manual debugging, and weak generalization ability. In this paper, a DE-Transformer-BiLSTM regression prediction algorithm is proposed, which combines the advantages of differential evolution optimization, Transformer global feature extraction and BiLSTM bidirectional timing modeling to automatically optimize key parameters and take into account both global correlation and local timing features. The experiment adopts 1283 effective monitoring data from industrial site, sets 10 input characteristics and boiler thermal efficiency as output variables, and carries out comparative experiments after data preprocessing and correlation analysis. The results show that the model has MSE of 0.142, RMSE of 0.377, MAE of 0.286, MAPE of 0.308 and R^2 of 0.93. All indexes are better than many traditional models, which can effectively improve the prediction accuracy and generalization ability of boiler thermal efficiency under complex working conditions, and provide a new method for energy efficiency prediction of active systems.

Keywords: Industrial Steam Boilers, Thermal efficiency prediction, Transformer, BiLSTM

1. Introduction

Industrial steam boiler is the core equipment of energy power system, which is widely used in electric power, chemical industry, metallurgy and other fields. Its thermal efficiency directly determines energy consumption, operating cost and carbon emission level [1]. Under the current energy transformation and dual-carbon goal, boilers have participated in deep peak shaving for a long time, and are frequently in complex working conditions such as variable load, coal quality fluctuation, equipment aging, etc. Traditional mechanism modeling is difficult to adapt to nonlinear, strongly coupled and time-varying thermal processes [2]. There are many and closely related boiler operating parameters, and dynamic changes such as exhaust loss, incomplete combustion loss and heat dissipation loss make accurate prediction of thermal efficiency a difficult problem in the industry [3]. Traditional counter-balance calculation relies on offline laboratory data, which has poor real-time performance. However, a single empirical model cannot cover the changes of all working

DOI: 10.54254/2755-2721/2026.33645

conditions, which is difficult to meet the dual requirements of high efficiency, energy saving and safe and stable operation in industrial sites. Building a high-precision and strong generalization thermal efficiency prediction model has important engineering value for optimizing combustion control, reducing energy consumption and improving the intelligence level of the system [4].

Machine learning provides a data-driven and effective path for boiler thermal efficiency prediction, breaking through the limitation of traditional methods that rely on mechanism assumptions and static formulas. The algorithm can autonomously mine the implicit relationship between load rate, exhaust temperature, excess air coefficient, coal quality parameters and thermal efficiency from massive operation data, and adapt to the high-dimensional nonlinear mapping relationship [5]. Traditional machine learning methods such as random forest and support vector machine can realize basic regression, but the accuracy is insufficient when dealing with timing dependence, long-range correlation and dynamic fluctuation. LSTM in deep learning can capture temporal local dependencies, and Transformer captures global features through self-attention mechanism, but it is difficult for a single model to balance global association and local timing details. The existing methods generally have problems such as parameter dependence on manual debugging, easy to fall into local optimum, weak generalization ability, etc. The prediction error is too large under variable working conditions and multivariable coupling scenarios, which can't meet the demand of industrial high-precision prediction.

Aiming at the shortcomings of existing models, this paper proposes a DE-Transformer-BiLSTM regression prediction algorithm, which combines the advantages of differential evolution optimization, Transformer global feature extraction and BiLSTM bidirectional timing modeling. The algorithm uses BiLSTM to bidirectionally capture the forward and reverse dependencies of boiler running sequence, and strengthen local dynamic feature learning; Transformer multi-head self-attention mechanism is used to mine the long-range global correlation between multi-parameters, which solves the problems of poor parallelism and long-sequence information decay in traditional recurrent neural networks. The differential evolution algorithm is introduced to automatically optimize the key parameters of the model, instead of manual parameter adjustment, and guide the model to jump out of the local optimum, thus improving the convergence efficiency and prediction stability. The model takes into account the dynamic characteristics of time series and the global coupling relationship, realizes multivariable collaborative modeling and parameter adaptive optimization, which can effectively improve the prediction accuracy and generalization ability of boiler thermal efficiency under complex working conditions, and provide an innovative method for energy efficiency prediction of active systems.

2. Data introduction

The data set in this paper is derived from the continuous operation monitoring data of industrial site. The original samples are obtained through real-time acquisition of distributed sensors and summary of distributed control system. After preprocessing processes such as outlier elimination, missing value filling, data smoothing and standardization, 1283 valid data are finally formed. The data set contains 10 input features and 1 output variable, specifically covering boiler load rate, feed water temperature, exhaust temperature, excess air coefficient, coal ash content, coal moisture, furnace negative pressure, steam pressure, combustion efficiency and continuous operation hours. The target variable is boiler thermal efficiency. First, a correlation analysis was performed, and a correlation heat map was plotted, as shown in Figure 1.

DOI: 10.54254/2755-2721/2026.33645

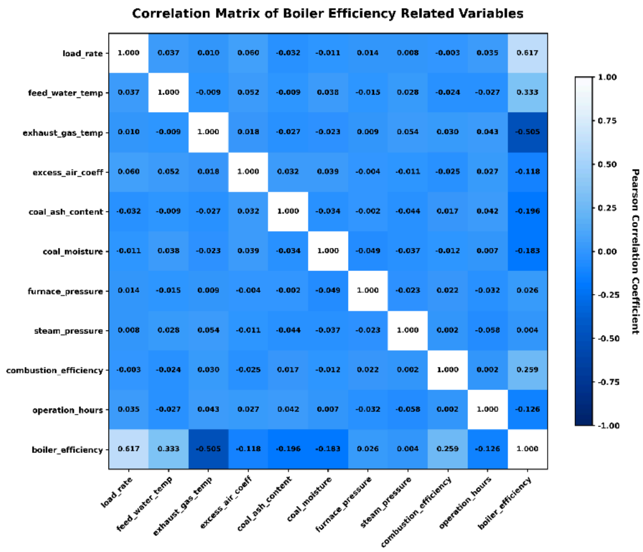

Figure 1. The correlation heat map

Plot a violin plot for the individual variables, as shown in Figure 2.

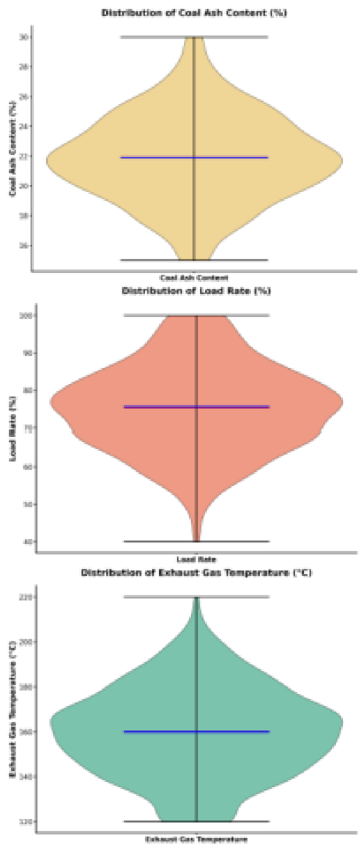

Figure 2. The violin plot

DOI: 10.54254/2755-2721/2026.33645

3. Methods

3.1. Differential evolution algorithm

Differential evolution algorithm is a kind of optimization algorithm based on swarm intelligence, which realizes global optimization by simulating biological evolution process. Initially, the algorithm generates a set of random solutions to form a population, and then superimposes the vector differences of different individuals on the target individual through mutation operation to generate mutated individuals. Then using crossover operation to exchange information between the mutant individual and the target individual, the test individual is obtained. Finally, the fitness of test individuals and target individuals is compared by selection operation, and the better individuals are retained to enter the next generation. The whole process continuously iterates, gradually approaching the optimal solution, and has strong global search ability and robustness [6]. The network structure of the differential evolution algorithm is shown in Figure 3.

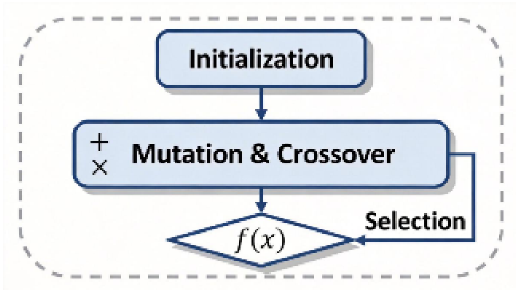

Figure 3. The network structure of the differential evolution algorithm

3.2. Transformer

Transformer algorithm is a deep learning model based on self-attention mechanism, which completely relies on attention mechanism to realize sequence information modeling. The model abandons the cyclic structure and captures the dependencies of elements at different positions in the sequence through multi-head self-attention, which can efficiently correlate the elements regardless of their distance [7]. At the same time, position coding is used to inject sequence sequence information to ensure that the model understands the timing characteristics. The encoder is responsible for extracting the deep features of the input sequence, while the decoder generates the target sequence based on the encoded information, supplemented by feedforward neural network and residual connection to improve the training stability [8]. The network structure of Transformer is shown in Figure 4.

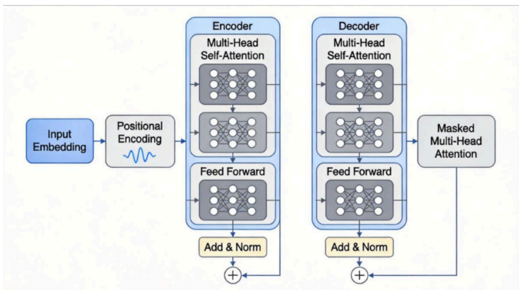

Figure 4. The network structure of Transformer

DOI: 10.54254/2755-2721/2026.33645

3.3. Two-way long-short-term memory network

Bidirectional long short-term memory network is an improved structure of recurrent neural networks, which can effectively solve the gradient disappearance problem of traditional recurrent networks. The network is composed of a plurality of memory units, each unit contains an input gate, a forget gate and an output gate, and controls the retention and transmission of information through gating mechanism. The model is divided into two independent LSTM layers, forward and backward. The forward layer reads the information in sequential order, and the backward layer processes the information in reverse order, and finally fuses the bidirectional features [9]. This structure can make use of the past and future context information simultaneously to capture the inherent laws of time series data more comprehensively. The network structure of the bidirectional long short-term memory network is shown in Figure 5.

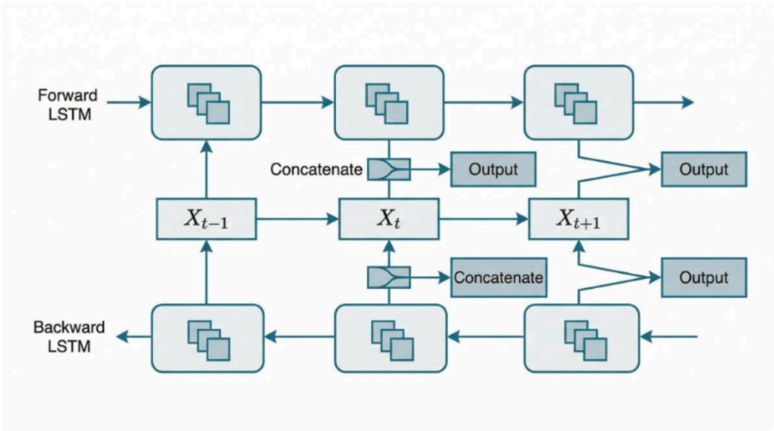

Figure 5. The network structure of the bidirectional long short-term memory network

3.4. DE-Transformer-BiLSTM

DE-Transformer-BiLSTM regression algorithm is a hybrid regression model that fuses optimization and deep learning, combining the advantages of differential evolution algorithm and two neural networks. Firstly, the differential evolution algorithm is used to optimize the key hyperparameters and initial weights of the model, so as to avoid the neural network training falling into local optimum and improve the overall convergence effect of the model. Then, the long-distance dependent features in the sequence are extracted through Transformer module, and the global association information is fully mined. Then the extracted features are input into the BiLSTM network to further capture the bidirectional timing dependence and local change law. The output layer of the network obtains the final regression result through linear transformation, which makes the model have both global optimization ability and accurate timing fitting ability, effectively improving the accuracy of regression prediction [10]. The network structure of DE-Transformer-BiLSTM is shown in Figure 6.

DOI: 10.54254/2755-2721/2026.33645

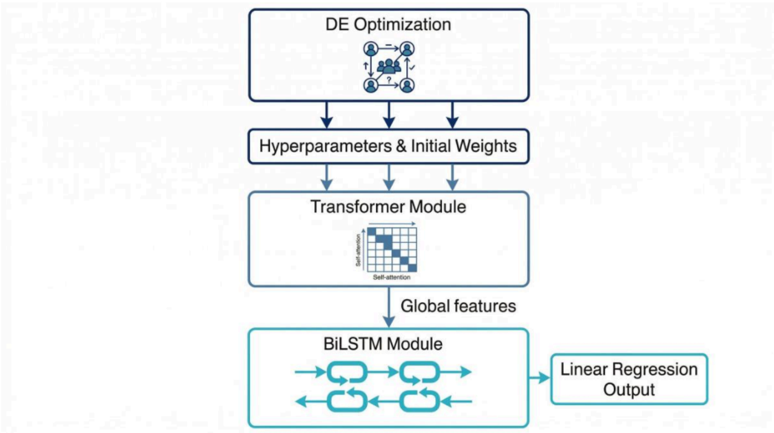

Figure 6. The network structure of DE-Transformer-BiLSTM

4. Results

The optimization algorithm of this project sets the population number to 8, the maximum number of iterations to 6, and the parameter dimension to be optimized to 3, where the lower bound of learning rate optimization is 0.001 and the upper bound is 0.01, the lower bound of the number of hidden layer nodes is 10 and the upper bound is 30, and the lower bound of L2 regularization coefficient is 0.0001 and the upper bound is 0.1. In the network model training stage, the maximum number of training rounds is set to 1000, the number of batches is set to 256, the learning rate decrease factor is set to 0.1, the learning rate decrease period is set to 850 times, the gradient clipping threshold is set to 10, the number of self-attention mechanism heads is set to 4, the total number of key channels is set to 128, the maximum position coding is set to 512, the discarded layer ratio after activation function is set to 0.01, and the proportion of training set in data set division is set to 0.7.

Output the comparative test results of each machine learning algorithm, as shown in Table 1. Output the index comparison histogram of the DE-Transformer-BiLSTM test set, as shown in Figure 7.

Table 1. Model evaluation parameters

Model	MSE	RMSE	MAE	MAPE	R^2
ExtraTrees	0.36	0.6	0.489	0.529	0.8
GBDT	0.262	0.512	0.408	0.442	0.856
CatBoost	0.279	0.528	0.424	0.459	0.864
Decision tree	0.817	0.904	0.716	0.774	0.558
BP neural network	0.319	0.565	0.434	0.468	0.841
Random forest	0.406	0.637	0.502	0.543	0.78
AdBoost	0.324	0.569	0.449	0.486	0.842
XGBoost	0.21	0.458	0.361	0.39	0.886
DE-Transformer-BiLSTM	0.142	0.377	0.286	0.308	0.93

DOI: 10.54254/2755-2721/2026.33645

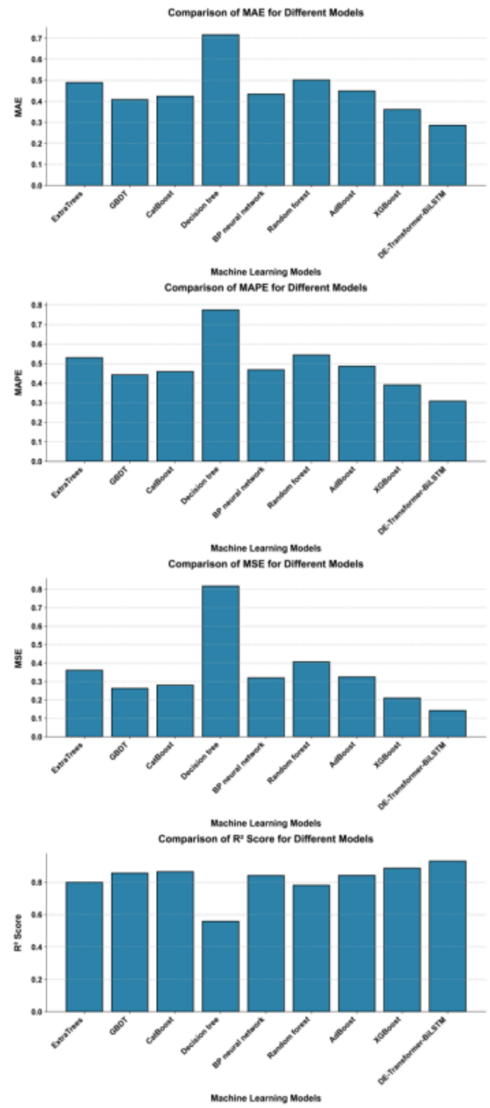

Figure 7. The index comparison histogram of the DE-Transformer-BiLSTM test set

The DE-Transformer-BiLSTM algorithm proposed in this paper is significantly better than ExtraTrees, GBDT, CatBoost, Decision Tree, BP neural network, random forest, AdBoost and XGBoost in various regression indicators. Its MSE, RMSE, MAE and MAPE are the lowest values among all models, and R^2 reaches the highest 0.93, fully demonstrating smaller prediction error and stronger data variance explanation ability. Compared with XGBoost, which performs suboptimally, this algorithm achieves further reduction in error indicators, and the fitting effect and prediction accuracy are ahead in all aspects. The decision tree performance of traditional machine learning models is the worst. Although various ensemble learning models and BP neural network are better than a single decision tree, the overall error level is still much higher than that of DE-Transformer-BiLSTM, highlighting the significant advantages of this hybrid optimization model in complex feature capture and timing law mining.

5. Conclusion

In order to solve the problem of accurate prediction of industrial steam boiler thermal efficiency, a DE-Transformer-BiLSTM hybrid prediction model is constructed to solve the problem of poor adaptability of traditional methods under nonlinear strong coupling time-varying conditions. The model realizes the automatic optimization of hyperparameters and initial weights with the help of differential evolution algorithm, avoids the defect of manual parameter adjustment and local optimal problem, and improves the convergence efficiency and stability of the model. Transformer multi-head self-attention mechanism is used to mine multi-parameter long-range global correlation, which makes up for the shortcomings of long-sequence information attenuation in traditional recurrent networks, and then bidirectionally captures timing dynamic dependence through BiLSTM to strengthen local change law learning. The experimental results show that the proposed model is optimal in MSE, RMSE, MAE, MAPE and other error indexes, with R^2 reaching 0.93. The prediction accuracy and fitting effect are significantly better than ExtraTrees, GBDT, CatBoost, XGBoost and other ensemble learning and traditional neural network models, and the achievement error is further reduced compared with the sub-optimal XGBoost model. The model can adapt to complex operation scenarios such as variable load and coal quality fluctuation of boilers, provide reliable support for boiler combustion optimization, energy consumption reduction and intelligent operation, and also provide technical ideas that can be used for reference for energy efficiency prediction and intelligent control of similar active equipment, and has good engineering application value and popularization prospect.

References

[1] Ross-Veitía, Bárbara D., et al. "Machine learning regression algorithms to predict emissions from steam boilers." Heliyon 10.5 (2024).
[2] Aslan, Emrah, Yıldırım Özüpak, and Feyyaz Alpsalaz. "Boiler efficiency and performance optimization in district heating and cooling systems with machine learning models." Journal of the Chinese Institute of Engineers 48.7 (2025): 1115-1130.
[3] Lianzhai, Duan, et al. "Machine learning algorithms for prediction of boiler steam production." International Journal of Advances in Data and Information Systems 5.2 (2024): 157-172.
[4] Blackburn, Landen D., et al. "Dynamic machine learning-based optimization algorithm to improve boiler efficiency." Journal of Process Control 120 (2022): 129-149.
[5] Kani, Ghaem Taghipour, and Amirreza Ghahremani. "Predicting the thermal performance of heat pipes applying various machine learning methods and a proposed correlation." International Communications in Heat and Mass Transfer 142 (2023): 106671.
[6] Qin, Hai, et al. "A data-driven fault prediction integrated design scheme based on ensemble learning for thermal boiler process." 2020 IEEE International Conference on Industrial Technology (ICIT). IEEE, 2020.
[7] Eskandari, Erfan, et al. "Enhanced predictive modeling of Nusselt number in boiler tubes: numerical simulations and machine learning for water and SiO2/water." Journal of Thermal Analysis and Calorimetry 149.17 (2024): 10119-10148.
[8] Strušnik, Dušan, Marjan Golob, and Jurij Avsec. "Artificial neural networking model for the prediction of high efficiency boiler steam generation and distribution." Simulation Modelling Practice and Theory 57 (2015): 58-70.
[9] Ye, Tuo, et al. "Modeling and optimization of the NOX generation characteristics of the coal-fired boiler based on interpretable machine learning algorithm." International Journal of Green Energy 19.5 (2022): 529-543.
[10] Gladilin, N. D., et al. "Development of a Model for Predicting the Load on a Steam Boiler of a Thermal Power Plant." Optical Memory and Neural Networks 34.Suppl 2 (2025): S185-S195.