

Research on Electric Heating Energy Consumption Prediction Based on SSA-LSTM-Attention

Yiting Chang^{1,a}, Suli Zhang^{2,*}, Guanghao You^{1,b}

¹School of Energy and Power Engineering, Changchun Institute of Technology, Changchun, China

²School of Computer Technology and Engineering, Changchun Institute of Technology, Changchun, China

^achangyiting@stu.ccit.edu.cn, ^byouguanghao@stu.ccit.edu.cn

*Corresponding author: zhangsl@ccit.edu.cn

Abstract: In response to the challenge of low energy prediction accuracy caused by complicated parameter settings and the stochastic nature of LSTM networks used in electric heating energy consumption prediction, this study introduces an enhanced LSTM attention model. This model combines the Sparrow Search Algorithm (SSA) with Long Short-Term Memory (LSTM) neural networks and incorporates an attention mechanism to optimize predictions of electric heating energy consumption. The experimental findings indicate that the SSA-LSTM-Attention model achieves notable advancements in both prediction efficiency and accuracy when contrasted with traditional baseline models like SVR, RF, BP, LSTM, and LSTM-Attention, especially when dealing with complex scenarios involving highly dynamic changes in electric heating data.

Keywords: Electric Heating; Energy Consumption Prediction; Long Short-term Memory Neural Network; Sparrow Search Algorithm

1. Introduction

The advent of carbon peak and carbon neutral targets has precipitated the gradual emergence of electricity substitution as a pivotal energy transition strategy on a global scale. Concurrently, the economic and environmental advantages of electricity as a clean energy source are becoming increasingly evident ^[1]. As set forth in the Guiding Opinions on Electric Energy Substitution issued by the National Energy Administration ^[2], residential coal-to-electricity conversion projects should be promoted in accordance with local conditions with the objective of reducing carbon dioxide emissions ^[3]. As a novel energy consumption paradigm, electric heating can be employed to regulate load, absorb new energy, mitigate the phenomenon of abandoned wind and light, and reduce carbon dioxide emissions. China's northern regions are actively promoting the "coal-to-electricity" program, encouraging users to utilize electric heating equipment for heating purposes. The aim is to enhance environmental efficiency, diminish reliance on conventional energy sources like coal, and advocate for the sustainable evolution of the energy industry infrastructure towards greener alternatives.

In the field of electric heating energy consumption prediction, meteorological factors exert a significant influence on short-term energy consumption prediction, which should not be overlooked. Electric heating energy consumption is a temperature-controlled load ^[4], and in recent years, scholars at home and abroad have conducted a number of studies on this topic. The studies focused predominantly on analyzing meteorological variables such as temperature, humidity, wind speed, and precipitation and proposed novel prediction strategies that diverged from traditional methods. The literature ^[5] examines the temporal characteristics of meteorological influences on urban power consumption and introduces a novel prediction approach. Nevertheless, it is possible that conventional weather forecasts and time-series convolutional networks may not represent the optimal solution. The literature ^[6] examines the impact of weather-sensitive load forecasting models, with a particular focus on meteorological factors such as temperature, humidity, and snow, and their influence on model formulation and data processing strategies. In the literature ^[7], a historical dataset is formed by analyzing the characteristics of electric heating loads, such as peaks, minima, time variations, fluctuations, and sudden changes. The KELM model is then used to train these data in order to achieve short-term load forecasting for electric heating. This effectively reduces the impact of random fluctuations on forecasting accuracy. However, the method does not take into account the potential impact of meteorological data. In contrast, Literature ^[8] employs

real-time meteorological data to establish summer weather-sensitive load models, exploring the relationships between factors like temperature, humidity, and wind speed through integrated regression analysis. However, this approach primarily considers the impact of temperature within specific forecasting periods, which may limit its ability to predict high-capacity load demands. Literature [9] employs the LSTM algorithm for electric heating load forecasting, adjusting step lengths to accommodate various load cycles. The simulation results validate the effectiveness of the method in meeting diverse forecasting requirements.

The present paper puts forth a methodology for short-term energy consumption prediction in electric heating based on the use of an SSA-LSTM-Attention model. The method employs an LSTM model to efficiently examine long-term dependencies in the data, discern patterns and trends, and apply them to the historical dataset. Concurrently, the attention mechanism is incorporated to enable the dynamic adjustment of attention at varying time points in accordance with the prevailing input data, thereby facilitating more precise pattern detection and information extraction from the time series and consequently enhancing the precision of the prediction. Subsequently, the prediction accuracy is enhanced by optimizing the hyper-parameter settings of the model through SSA. The feasibility and effectiveness of the method are then verified by actual data.

2. SSA-LSTM-Attention Based Energy Consumption Prediction for Electric Heating

2.1. Sparrow search algorithm

A novel population optimization algorithm, designated the SSA algorithm, has been developed by researchers inspired by the foraging and anti-predatory behaviors observed in sparrows. The algorithm is employed extensively for parameter optimization in prediction models due to its exceptional performance in terms of search accuracy and convergence speed. In a sparrow population engaged in foraging activities, two distinct categories of sparrows can be identified: discoverers and joiners. The position of the explorer is constantly changing as he explores and encounters predators, and the position is updated in the manner shown in equation (1):

$$X_{ij}^{i+1} = \begin{cases} X_{ij}^i \exp\left(-\frac{1}{\alpha I_{max}}\right), & R_2 < S \\ X_{ij}^i + QL, & R_2 \geq S \end{cases} \quad (1)$$

Where $X_{i,j}$ represents the spatial coordinates of the i th sparrow along the j th dimension; I_{max} denotes the maximum number of iterations; α is a random number taking values in the range of $(0,1]$ $R_2 \in [0,1]$ and $S \in [0.5, 1]$, represents the warning and safety values, respectively; Q follows a normal distribution; and L is a row matrix of size $1 \times d$ with all elements set to 1.

The position of the joiner is updated in accordance with the following equation (2):

$$X_{ij}^{i+1} = \begin{cases} Q \exp\left(\frac{X_{worst}^i - X_{ij}^i}{i^2}\right), & i > \frac{n}{2} \\ X_p^{i+1} + |X_{ij}^i - X_p^{i+1}| A^+ L, & \text{other} \end{cases} \quad (2)$$

Where X_p represents the best foraging position found by the forager; X_{worst} represents the current global worst position; A is a matrix size $1 \times d$ whose elements are randomly taken from $\{-1,1\}$, and with $A^+ = A^T(AA^T)^{-1}$; n is the size of the population.

Sparrow populations are warned when a predator is detected and their location is updated as in equation (3):

$$X_{ij}^{i+1} = \begin{cases} X_{best}^i + \beta |X_{ij}^i - X_{best}^i|, & f_i > f_g \\ X_{ij}^i + K \frac{|X_{ij}^i - X_{worst}^i|}{f_i - f_g + \varepsilon}, & f_i = f_g \end{cases} \quad (3)$$

Where X_{best} denotes the optimal position across the entire sparrow population; β serves as a parameter for controlling step size. k indicates the direction of sparrow movement, also acting as a step control parameter. f_i and f_g indicates the direction of sparrow movement, also acting as a step control parameter. ε is a constant employed to prevent division by zero in calculations.

2.2. Electric Heating Energy Consumption Prediction Based on LSTM-Attention

LSTM is based on the traditional RNN and introduces two key concepts: the cell state and the gating mechanism. The cell state enables the continuous transmission of relevant sequence information, while the gating mechanism allows for the accurate control of the addition and elimination of information. These two concepts endow the model with the ability to retain information over long or short periods of time, thus providing a solution to the limitations of traditional RNN models. This design effectively mitigates the challenges of gradient explosion and vanishing gradients commonly encountered in RNN. The architecture of LSTM is depicted in Figure 1 below:

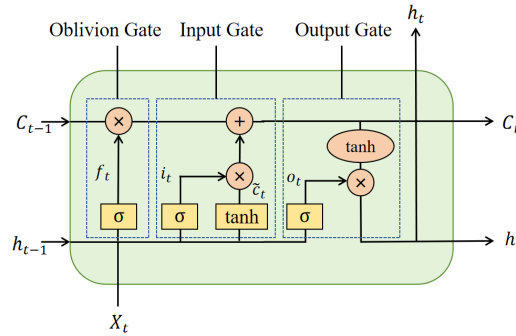


Figure 1: Network structure diagram of LSTM.

The gating mechanism and the calculation of the cell state are shown in equations (4)-(9):

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (6)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (7)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \tanh(C_t) \quad (9)$$

Where W represents the weight matrix of the gating mechanism; b denotes the bias vector of the gating mechanism; and f_t , i_t , and o_t are the outputs of the forgetting gate, input gate and output gate respectively. The forgetting gate decides whether to retain or discard information, the input gate selectively modifies the cell state, and the output gate determines the value of the subsequent hidden state. The input at the current moment x_t is represented by the variable. The cell state at the previous moment C_{t-1} , and the cell state at the current moment C_t are respectively the cell states processed by the gating mechanism in the current moment and the subsequent moment.; Similarly, the hidden states at the previous moment h_{t-1} and the hidden states at the current moment, h_t are respectively the hidden states processed by the gating mechanism in the current moment and the subsequent moment. Finally, σ denotes the sigmoid function, while \tanh denotes the hyperbolic tangent function.

The conventional approach to neural networks is to treat all inputs with equal weight, despite the potential for some inputs to be more significant than others. In practical engineering applications, such as forecasting energy consumption for electric heating, information pertaining to weather and date type may be of greater consequence than other inputs. Attention mechanisms are employed to accentuate the influence of these critical inputs on the outputs, thereby enhancing the precision and efficacy of the predictions. Figure 2 below illustrates a common structure for an attention mechanism.

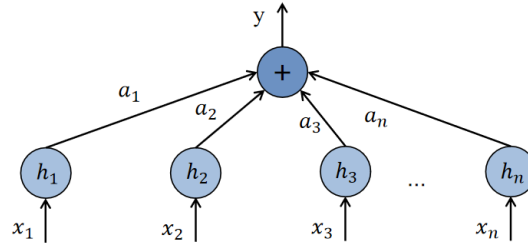


Figure 2: Common attention mechanism structure diagram.

The attention mechanism enhances the precision of the network by assigning distinct weights to disparate inputs, thereby enabling the neural network to concentrate more on significant information. In the attention mechanism, the inputs are divided into multiple subspaces, each with unique weights that can be learned or predetermined based on the specific requirements of the problem at hand. In each computational step, the weights associated with each subspace are adjusted in a dynamic manner, with the objective of aligning them with the specific requirements of the task at hand. Figure 2 depicts the structure of the attention mechanism, where x_i represents the input, h_i represents the output value of the LSTM, and a_n represents the value of the attention weight of the attention mechanism on the output layer of the LSTM.

The traditional LSTM neural network demonstrates robust predictive capacity when confronted with time-series data. However, it is susceptible to the phenomenon of early learning content being forgotten for long series samples, which may result in the loss of crucial information and, consequently, impair the accuracy of the prediction. Accordingly, the attention mechanism has been incorporated into the prediction model to accentuate the salient features that influence the energy consumption of electric heating, thereby enhancing the model's predictive capacity. The model structure is illustrated in Figure 3, showing the input layer, LSTM layer, attention mechanism layer, and output layer.

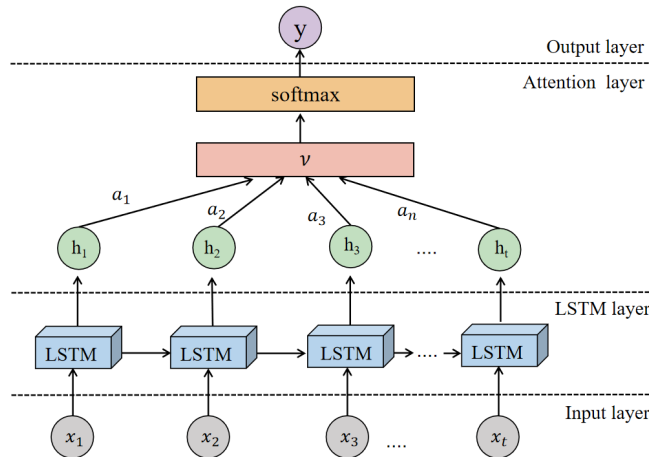


Figure 3: Common attention mechanism structure diagram.

The input layer of the model receives the fluctuation term influence factors associated with the energy consumption data x_t and transmits them to the LSTM neural network. In the LSTM layer, the sample features are learned to generate the corresponding hidden layer output h_t . Subsequently, an attention mechanism layer is introduced, which serves the purpose of dynamically adjusting the weights in accordance with the input features and continuously updating these weights. These weights are calculated using equations (10)-(12).

$$e_t = \tanh(\omega_1 h_t + b) \quad (10)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^t e_j} \quad (11)$$

$$v_t = \sum_{i=1}^t \alpha_i h_i \quad (12)$$

where the hidden layer vector e_t represents the hidden state at moment t ; α_t is the weights assigned by the attention mechanism to the different input features; the computation of these weights involves the weight coefficients u and ω_1 , as well as the output v_t of the attention mechanism layer; and the effect of the bias vector b is also included.

The final stage of the model is the output layer, which transforms the input data into a predicted result for the subsequent moment. This is achieved through a fully connected layer, as illustrated by equation (13).

$$y_t = \sigma(\omega_2 v_t + b) \quad (13)$$

Where y_t is the predicted output value; ω_2 is the weight matrix, which determines the ability of the model to predict the output given the input. b is the bias vector; Let σ denote the Sigmoid function.

2.3. Electrical Heating Energy Consumption Prediction Process

Figure 4 illustrates the flow chart for predicting electric heating energy consumption:

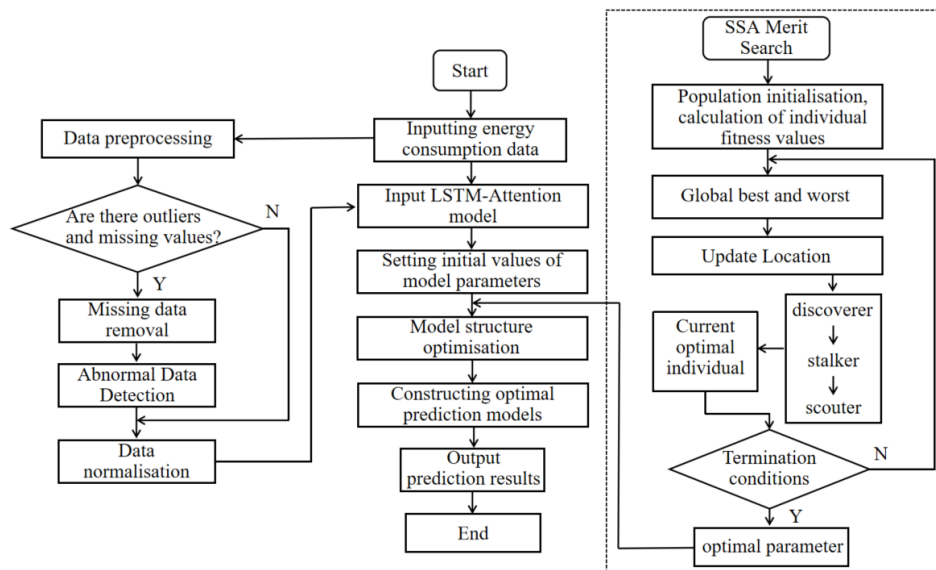


Figure 4: Flowchart of electric heating energy consumption prediction.

(1) Initially, the data pertaining to electric heating energy consumption are subjected to preprocessing to ascertain the presence of outliers and missing values. In the event of such anomalies, anomaly detection and processing are conducted to guarantee the integrity and veracity of the data. In the absence of anomalies, normalization is performed directly, and the normalized data are employed as the input variables of the LSTM-Attention model.

(2) The initialization of the parameters of the SSA is to be carried out; the key parameters of the LSTM-Attention model are to be mapped to the individuals in the SSA algorithm; the error is to be calculated through these parameters in order to assess their fitness; and the speed and position of the sparrows are to be adjusted according to the current optimal position in order to form a new population of individuals. A further prediction is made by the LSTM-Attention model using the new parameters, and the resulting error and fitness values are calculated. In the event that the new fitness value is superior to the existing optimal solution, the optimal solution and its corresponding position are to be updated; otherwise, the optimal solution is to remain unchanged. These steps are to be repeated continuously in order to update the position and speed of the sparrows, generate new groups of sparrows, and evaluate their fitness until the termination condition is satisfied.

(3) Ultimately, the model parameters that correspond to the optimal fitness value are output and applied to optimize the structure of the LSTM-Attention model, thereby facilitating the construction of an optimal prediction model for the prediction of electric heating energy consumption. This process not only ensures the accuracy of the predicted outcome but also optimizes the performance of the model.

3. Experimental Analysis

In this study, simulation tests were conducted using electric heating energy consumption data, meteorological data, and date classification in a region of Jilin Province during the heating period from 2021 to 2023. The data set comprised 769 sets of data, including energy consumption, maximum air temperature, minimum air temperature, average air temperature, weather, wind, body temperature, and date type. The data was divided into a training set and a validation set in an 8:2 ratio. In order to ascertain the veracity of the proposed model, five alternative models, namely BP, SVR, LSTM, SSA-LSTM, and LSTM-Attention, have been constructed for comparative analysis.

3.1. Experimental Platform

This experiment is based on Python language and compiled using Jupyter Notebook, the specific experiment configuration is shown in Table 1 below:

Table 1: Experimental configuration table.

Name	Version
Compiled Language	Python3.9
Compilation Tools	Jupyter Notebook
Deep Learning Framework	TensorFlow 2.10.0
Operating System	Windows 11
Processor	11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz
Hard Disk Capacity	155GB+320GB

3.2. Data Preprocessing

The collection process of electric heating energy consumption data may be affected by equipment anomalies, resulting in the omission of data or the inclusion of anomalous data. To address this quality issue, it is essential to examine and eliminate any anomalous data and subsequently replace it with a value of zero. In the case of missing data, two potential solutions exist: the first is to fill the gap with a value of zero as an independent feature, while the second involves utilizing the mean value of the data preceding and succeeding the missing data point. In the event of a significant and continuous absence of data, it may be appropriate to utilize the mean value of the complete data set as a substitute. In the event of a minor data deficit, the most expedient solution is to remove the affected data directly, thereby reducing the noise inherent to the model. However, this approach may also result in the disruption of the data's overall regularity. Accordingly, in this paper, the strategy entails the utilization of the mean value of the preceding and subsequent weeks in order to fill in the anomalous data. Figure 5 shows a comparison between the abnormal and normal energy consumption data.

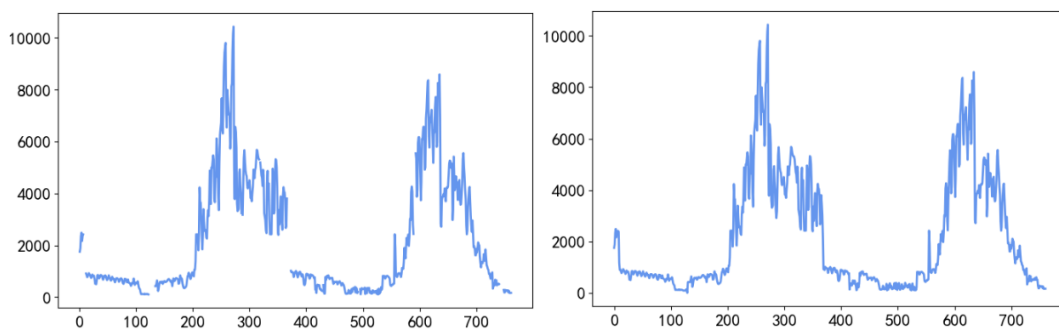


Figure 5: Comparison of abnormal data and normal data of energy consumption.

3.3. Correlation Analysis

Furthermore, in order to conduct a comprehensive examination of the correlation between electric heating energy consumption and air temperature, relative humidity, and holidays, a correlation analysis can be performed utilizing the Pearson correlation coefficient. The dataset employed in this experiment encompasses 762 data points, encompassing the variables under investigation across a range of time periods. In the analysis, weekdays are represented by the number 1, and non-working days are represented by the number 0. The resulting Pearson's correlation coefficient analysis histogram is

presented in the figure below. Figure 6 illustrates a significant correlation between the three variables of maximum temperature, minimum temperature, and average temperature and energy consumption. These variables effectively explain the other factors.

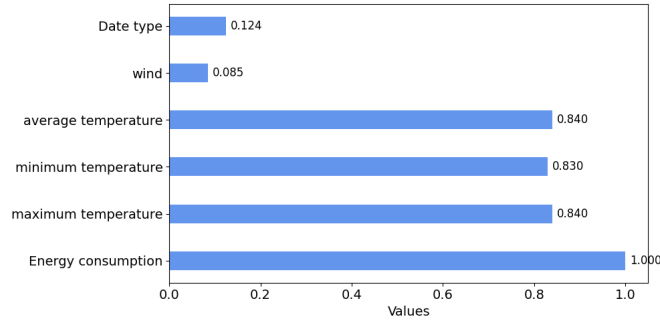


Figure 6: Bar chart of Pearson correlation coefficient analysis.

3.4. Evaluation Index

In order to verify the effectiveness of the proposed prediction model for electric heating energy consumption data, three evaluation indices are used in this section to measure its predictive effectiveness, including the root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2). The expressions of the evaluation indices are given in equations (14)–(16). RMSE is a statistical measure that quantifies the average squared difference between the predicted value and the actual value. It is a reflection of the magnitude of the prediction error. MAE represents the mean of the absolute differences between the predicted and actual values and is employed to assess the average absolute magnitude of the prediction error. R^2 measures the extent to which a model fits the data, or the amount of variance that the model explains. It ranges from 0 to 1, with values closer to 1 indicating a superior model fit.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (14)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (15)$$

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (16)$$

Where, n is the number of samples, Y_i is the i th actual value, \hat{Y}_i is the i th predicted value, and \bar{Y} is the average of the actual values.

3.5. Analysis of Experimental Results

In order to predict electric heating energy consumption, the LSTM-Attention model and the SSA-LSTM-Attention model, as proposed in this paper, were compared. Firstly, the parameters of the LSTM network were optimized by SSA, resulting in a learning rate of 0.0143, a dropout rate of 0.188, and a number of neural units in the hidden layer of 100 and 88, respectively. The experimental result graphs demonstrate that the SSA-optimized LSTM-Attention model exhibits markedly superior prediction ability compared to the original LSTM-Attention model. The capacity of the two models to fit the dataset is demonstrated in Figure 7 below. It can be observed that the optimized model is able to fit the original data with greater accuracy and predict the electric heating energy consumption curve with greater precision.

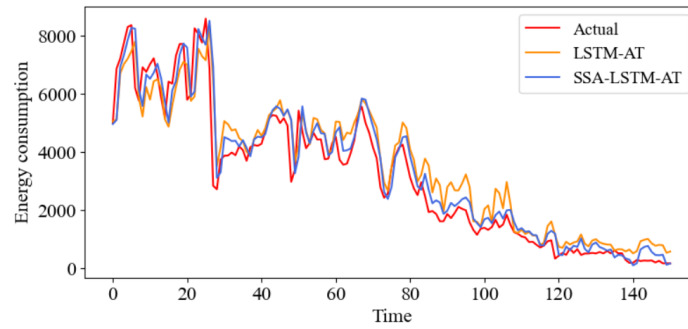


Figure 7: Comparison of electric heating energy consumption prediction results.

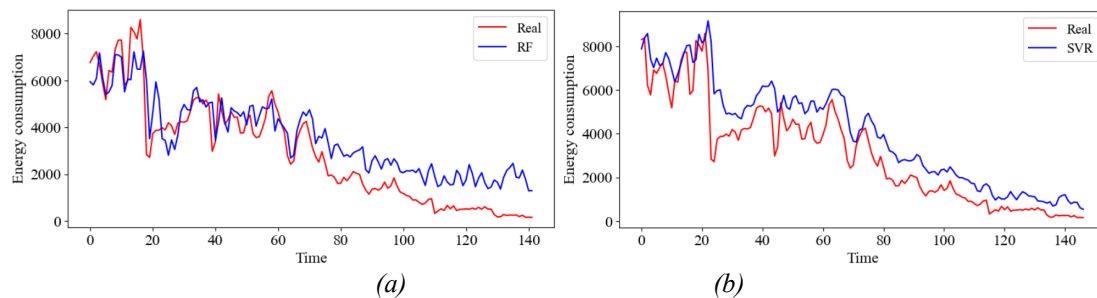
To further validate the model performance, the remaining four baseline models were selected for comparison with the SSA-LSTM-Attention model. These were BP, RF, SVR, and LSTM, as well as LSTM-Attention. The results of the error evaluation metrics for these six models are summarized in Table 2.

Table 2: Summary table of error evaluation index results for the six models.

Models	RMSE	MAE	R ²
SVR	1172.53	965.95	0.74
RF	1097.14	947.18	0.75
BP	1080.04	959.88	0.79
LSTM	990.95	844.85	0.82
LSTM-Attention	798.82	595.49	0.88
SSA-LSTM-Attention	678.03	441.06	0.91

As illustrated in Table 2, the SSA-LSTM-Attention model exhibits superior performance in terms of lower error and higher prediction accuracy on the experimental dataset. In particular, the RMSE of the SSA-LSTM-Attention model is 678.03, the MAE is 441.06, and the R² reaches 0.91. The reason for achieving this prediction result is that the introduction of SSA into the LSTM-Attention model facilitates optimization of the model structure and training strategy, thereby accelerating convergence and avoiding overfitting. This enhances the model's generalization ability and prediction accuracy. The traditional BP, RF, SVR, and single LSTM models are unable to effectively capture long-term dependencies and contextual information when dealing with the effects of climatic factors and holidays in the prediction of electric heating energy consumption due to their limited ability to capture long-term dependencies in complex time-series data. In particular, LSTM models may encounter difficulties in retaining information over extended periods or in the presence of complex periodic effects, leading to potential issues in their performance. In contrast, the SSA-LSTM-Attention model is better equipped to handle these intricate dependencies by incorporating SSA optimization and multi-level attention mechanisms, thereby enhancing prediction accuracy and the model's practical utility.

To further demonstrate the effectiveness of the model prediction, the test set data from the dataset is selected for visualization and analysis. Figure 8 shows the visualization of the prediction results. The graphs in (a-e) represent the prediction effects of the BP, SVR, RF, LSTM, and LSTM-Attention models on the dataset, respectively. It can be observed that, in comparison to the other baseline models, which exhibit errors at certain data points, the SSA-LSTM-Attention model demonstrates accuracy at the majority of the points when tested on the test set. This indicates that the model is capable of accurately predicting the energy consumption of electric heating.



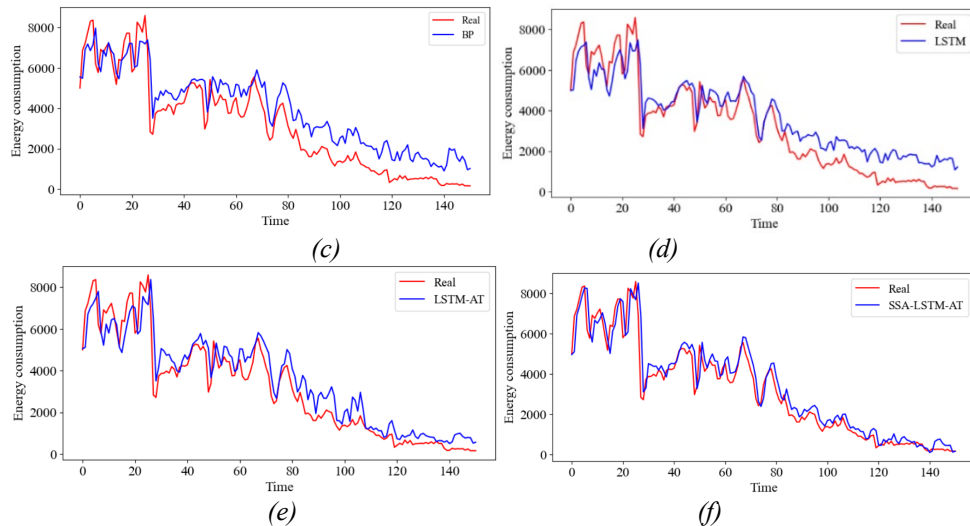


Figure 8: Comparison of the prediction effects of each model, including: (a)RF; (b)SVR; (c)BP; (d)LSTM; (e)LSTM-AT; (f)SSA-LSTM-AT.

4. Conclusion

This study proposes a novel electric heating energy consumption prediction model, namely SSA-LSTM-Attention, which integrates the sparrow search algorithm and a long- and short-term memory neural network with the introduction of the attention mechanism. In comparison with traditional prediction methods, including BP, SVR, and RF, as well as a single LSTM model and other combined models (LSTM-Attention), the SSA-LSTM-Attention model demonstrates superior accuracy in predicting the energy consumption of electric heating in a region of Jilin Province during the heating period from 2021 to 2023. Future research will concentrate on enhancing the real-time and dynamic adaptability of the model, including the investigation of methods for updating the model in a real-time data streaming environment and the integration of additional environmental data sources while accounting for more complex external influences. Furthermore, efforts will be directed towards enhancing the model's interpretability and generalisability.

References

- [1] Mei Shengwei, Guo Yongqing, Si Yang, et al. Solar thermal air flow power generation technology and its application prospects [J]. *High Voltage Technology*, 2019, 45(11):3401-3412.
- [2] Guidance on Promoting Electric Energy Substitution [C]// Nuclear Industry Survey and Design (No.2, 2016, Total No.93). National Development and Reform Commission; National Energy Administration; Ministry of Finance; Ministry of Environmental Protection; Ministry of Housing and Urban-Rural Development; Ministry of Industry and Information Technology; Ministry of Transport; Civil Aviation Administration of China; , 2016:4.
- [3] Li Xianglong, Zhang Baoqun, Ma Longfei, et al. Bayesian Electric Heating Load Prediction Based on Hidden data Information Mining [J]. *Electrical Measurement & Instrumentation*, 2018,55(24):88-93.
- [4] Huang Yafeng, Zhu Yujie, Mu Gang, et al. Evaluation of adjustable capacity of household heating load based on temperature forecast [J]. *Power Grid Technology*, 2018, 42(08):2487-2493.
- [5] Wei Xiaochuan, Wang Xingang. Big based on meteorological data of urban power load prediction [J]. *Electric measurement and instrument*, 2021, 58 (02): 90-95.
- [6] Guo Zhanwu, Zhang Zeya, Zhou Xinghua, et al. Meteorological factors of electric heating load prediction research [J]. *Electric measurement and instrumentation*, 2022, 59 (02): 154-158.
- [7] Chen Qifang, Xia Mingchao, Guo Min, et al. Based on data mining electric heating area of short-term load forecasting method [J]. *Power supply*, 2019, 4 (12): 48-54.
- [8] Shi Jiaqi, Zhang Jianhua. Load Forecasting Method Based on Multi-model Fusion Stacking Ensemble Learning Method [J]. *Proceedings of the CSEE*, 2019, 39(14):4032-4042.
- [9] Li Xianglong, Ma Longfei, Zhao Xiangyang, et al. Time scales based on LSTM networks [J]. *Electric heating load forecasting of power system and its automation*, 2021 (4): 71-75.