

Building energy consumption prediction based on VMD-CNN-BILSTM-Attention

Ruoheng Liu ^{1, #}, Le Huang ^{2, *, #}, Chenfu Yuan ^{3, #}

¹ School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou, China

² College of Energy and Mechanical Engineering, Shanghai University of Electric Power, Shanghai, China

³ School of architecture, Changzhou University, Changzhou, China

* Corresponding author: 894104249@qq.com

#These authors contributed equally.

Abstract. Accurate building energy consumption prediction can help buildings optimize energy allocation and save energy. To achieve accurate and reliable building energy consumption prediction, this paper proposes a building energy consumption prediction method based on VMD decomposition and noise reduction, i.e., CNN-BILSTM-Attention hybrid prediction. In this paper, we first use the VMD method to decompose the historical energy consumption series to obtain relatively smooth multiple components, which lays the foundation for accurate prediction. Then we combine Convolutional Neural Network (CNN) and Bi-directional Long and Short Term Memory (BILSTM) network to fully extract the Spatio-temporal characteristics of energy consumption data and then introduce the Attention mechanism to automatically assign corresponding weights to the BILSTM hidden layer states to distinguish the importance of different influencing factors on energy consumption and improve the prediction accuracy. To verify the effectiveness of the model, the most later analyzed with real building energy consumption data of a region, and the prediction results were shown by python, and the prediction results were improved by 40% compared with the traditional machine learning algorithm.

Keywords: Energy consumption prediction, convolutional neural network, bidirectional long- and short-term memory network, attention mechanism, variational modal decomposition.

1. Introduction

Since the twenty-first century, with global industrialization and urbanization, the world's climate has deteriorated and the problems of environmental pollution and energy consumption have become progressively more serious [1]. World energy consumption is increasing every year [2]. Energy consumption in buildings has become the largest energy consumption in the world [3]. Among the various energy sectors, buildings account for as much as 20.1%-40% of the total global energy consumption [4]. Among them, public buildings are particularly high in energy consumption, which also means that they have a large potential for energy savings, a feature that is considered to be a major highlight in building energy efficiency. [5]Public building energy consumption forecasting can help this paper to reasonably predict energy consumption and play a positive role in the sustainable development of buildings. Therefore, public energy consumption prediction has become a hot topic.

Many scholars have studied many methods for energy consumption prediction in the past, mainly time series analysis, machine learning, and deep learning3 species. Among them, the time series analysis method uses a linear approach that takes into account the time-series characteristics of the corresponding data and is suitable for predictions with low volatility, for example, Divina et.al [6], Cao and Yang [7] both used time series analysis methods to predict the energy consumption, and Kan [8] used a SARIMA-GARCH model to predict the energy consumption of air conditioning for an office building. In addition, machine learning models have been applied to various types of energy consumption forecasting, addressing the shortcomings of time series analysis in dealing with volatile and large data. These models can learn the nonlinear relationship between energy consumption and

various influencing factors, for example, Jozi et al. [9] and Cheng et al. [10] used many methods such as support vector machine, BP neural network, random forest, and gradient boosting machine, to predict the energy consumption of buildings. Huang [11] used four single machine learning algorithms, namely, multiple linear regression algorithm, extreme learning machine algorithm, extreme gradient descent algorithm and support vector regression, to build energy consumption prediction models to predict residential building energy consumption respectively. However, the machine learning models are inefficient and depend on data. BILSTM network, a two-way long short-term memory neural network can consider both the temporal and nonlinear nature of energy consumption data solves the problems of the above existing models. Wang [12] proposed a combined model of allometric CNN and BILSTM for electricity load prediction and demonstrated its superiority over the single-structured LSTM model and the combined CNN-LSTM model. The attentional mechanism boosts sequence learning tasks, making it a hot research topic in many fields. For example, Shao et al. [13] proposed a building energy consumption prediction model combining attention mechanism and LSTM and demonstrated its superiority. In addition, when dealing with energy consumption prediction problems, some algorithms are usually used to reduce the data impact in order to reduce noise, among which VMD algorithm is used because of its advantages such as better accuracy of complex data decomposition and better resistance to noise interference, for example, Xu et al. [14] used an integrated learning method of VMD to predict the energy consumption of ground source heat pump system (GSHP), Ruan et al. [15] proposed a hybrid VMD-based long and short-term memory neural network power consumption prediction model for building energy consumption and confirmed the effectiveness of VMD in improving accuracy by comparison, Zhao et al. [16] proposed a proposed method for predicting energy consumption of polysilicon production based on VMD-ASSA-LSSVM model using Variational Mode Decomposition (VMD) to decompose the energy consumption series into different characteristic scale energy consumption components and reduce the non-smoothness and complexity of the energy consumption series.

In this paper, we propose a VMD-CNN-BILSTM-Attention based method for power energy consumption prediction. Among them, CNN, a convolutional neural network, has an excellent performance in feature extraction of complex information and can effectively extract features affecting power energy consumption data; the BILSTM layer, a bidirectional long and short-term memory network, is used to extract bidirectional temporal features of sequence data; the features generated by the BILSTM hidden layer are used as the input of the Attention mechanism, and the attention mechanism is used to The features generated by the BILSTM hidden layer are used as input to the Attention mechanism, and the information extracted by the BILSTM layer is differentiated in importance to enhance the model's emphasis on key features; the VMD has better accuracy in decomposing complex data, and a non-recursive processing strategy is used for noise reduction, and finally, real energy consumption data is used as an example for prediction.

This paper is structured as follows, and in Section describes the VMD-CNN-BILSTM-Attention model used in detail. In Section III, the electrical energy consumption data of an elementary school in Shirley, NY, USA, in 2012 is predicted based on data provided by the EnerNOC public database to demonstrate the merits of the proposed framework. In section 4, the methodology used and the conclusions drawn are summarized and future research directions are proposed.

2. Model Introduction

2.1. Convolutional neural network

Convolutional neural networks, abbreviated as CNN, are neural networks developed for processing two-dimensional image data. CNNs can be very effective in automatically extracting and learning features from one-dimensional data sequences.

Convolutional neural network is an improvement on BP neural network, similar to BP, both use forward propagation to calculate output values and backward propagation to adjust weights and biases; the biggest difference between CNN and standard BP is that the neural units between adjacent layers

in CNN are not used to connect with all neural units like BP, but partially connected. In this paper, we will compare the CNN with the BP model in Chapter 3 to derive the advantages of the model.

The CNN model in this paper computes activation values by forward propagation, setting up a convolutional layer and a pooling layer. The convolutional layer uses 64 feature maps (filter=64) and reads the input sequence with a kernel size of 3-time steps (kernel size=3). The maximum pooling layer (pool size=1) simplifies the feature maps by sampling the original map layer.

$$Z(u, v) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} x_{i+u, j+u} \bullet k_{roti, j} \bullet x(i, j) \quad (1)$$

$$x(i, j) = 1, 0 \leq i, j \leq n$$

$$0, \text{others}$$

$$Z_{u,v}^{(l)} = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} x_{i+u, j+u}^{(l-1)} * k_{roti, j}^{(l)} * x(i, j) + b^{(l)} \quad (2)$$

$$x(i, j) = 1, 0 \leq i, j \leq n$$

$$0, \text{others}$$

$$a_{(u,v)}^{(l)} = f(z_{u,v}^{(l)})$$

$$Z_{i,j}^{(l+1)} = \beta^{(l+1)} \sum_{u=ir}^{(i+1)r-1} \sum_{v=jr}^{(j+1)r-1} a_{u,v}^{(l)} + b^{(l)} \quad (3)$$

$$a_{u,v}^{(l)} = f(z_{u,v}^{(l)})$$

(1) Is the RMS convolution layer definition equation

(2) Is the output expression of the convolution layer

(3) Is the output expression of the pooling layer

2.2. BILSTM

BiLSTM is a combination of a forward LSTM and a backward LSTM. The LSTM model consists of an RNN with input gates, forgetting gates, united states, and output gates. The LSTM model is based on the RNN with the addition of gates and memory units, which effectively prevents gradient explosion and gradient disappearance.

In this paper, the BiLSTM model is chosen to better predict the energy structure, the CNN processes part of the data, and the one-dimensional results are imported into the LSTM.

As shown in equation (4). First, the forgetting gate F reads the previous moment output $ht-1$ and the current input X_t , and the Sigmoid function determines the information that needs to be discarded to achieve the discard information function in the cell state.

$$F_t = \sigma(W_f * [ht - 1, xt] + b_f) \quad (4)$$

Second, the input gate I contains two processes, as in Eqs. (5) to (7), to both save the information to be memorized to the cell state and to add the updated information to the cell state.

$$I_t = \sigma(W_i * [ht - 1, xt] + b_i) \quad (5)$$

$$C_t = \tanh(WC * [ht - 1, xt] + b_c) \quad (6)$$

$$Ct = Ft * Ct - 1 + It * Ct \quad (7)$$

Finally, the output gate O is used to control how much information from the internal memory cell Ct is output to the hidden state ht, which is calculated as shown in equation (8).

$$Ot = \sigma(Wo * [ht - 1, xt] + bo) \quad (8)$$

$$ht = Ot * \tanh(Ct) \quad (9)$$

The forward propagation is shown in equations (4) to (9), and the output values of the neurons are calculated. The backward propagation is similar to the traditional feed-forward neural network, and the error back propagation algorithm is used. The error of the output layer is used to solve for each layer's weights and the weight coefficients update is completed by gradient descent.

2.3. Attention mechanism

The attention mechanism is derived from the characteristics of human vision for information processing. Human vision always tends to pay attention to relatively important information when processing information, and the Attention mechanism mimics this feature by assigning the size of weights acting to the importance of information, the more important the information is, the more weights are assigned, which in turn can improve the efficiency of information processing. the Attention mechanism solves the problem of information loss caused by too long sequences in LSTM by focusing on important information related to energy consumption, such as temperature, humidity and other weather factors. The structure of the Attention unit is shown in Fig.

The attentional state transition process is as in Eq.

$$s(h_t, \bar{h}_s) = h_t^T * W * \bar{h}_s \quad (10)$$

$$\alpha_{ts} = \frac{\exp(s(h_t, \bar{h}_s))}{\sum_{s=1}^S \exp(s(h_t, \bar{h}_{s'}))} \quad (11)$$

$$c_t = \sum_s \alpha_{ts} \bar{h}_s \quad (12)$$

$$a_t = f(c_t, h_t) = \tanh(W_c [c_t; h_t]) \quad (13)$$

Where α_{ts} is the attention weight value, h_t represents the hidden layer state value, and W is a matrix to be learned and the same matrix is used throughout the process. a_t is the final feature vector.

2.4. VMD

(1) Overview of the VMD model

VMD is a completely non-recursive modal variation and signal processing method that solves the limitations of EMD, which is sensitive to noise and sampled signals. VMD can determine the number of modal decompositions and adaptively search for each model's optimal center frequency and finite bandwidth to achieve the effective separation of intrinsic mode functions (IMF) and signal frequency domain division.

(2) VMD specific implementation steps

The main idea in the VMD algorithm lies in constructing and solving the variational problem, and the principle is shown in Eqs. (14)-(21) are shown in the following steps:

Step 1: Constructing the variational problem

$$x(t) = \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] \exp(-j\omega_k t) \quad (14)$$

$$\min_{\{u_k\}, \{\omega_k\}} \{ \sum_k \|\partial_t x(t)\|_2^2 \} \quad (15)$$

$$s. t. \sum_{k=1}^K u_k(t) = f \quad (16)$$

Eqs. (14) and (15) are multivariate functions of, and Eq. (16) is the constraint of each mode.

where f is the initial signal, K is the number of decomposed modes, k is the k th mode component and the center frequency after decomposition, t is the Dirac function, and $*$ is the convolution operation.

Step 2: Solving the variational problem

Then, the originally constructed constrained variational problem is cleverly transformed into an unconstrained variational problem by introducing the augmented Lagrangian function and combining the quadratic penalty term as well as the advantages of the Lagrangian multiplier method to obtain the augmented Lagrangian expression. As shown in Eqs. (17) and (18).

$$g(t) = \|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \quad (17)$$

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \|\partial_t x(t)\|_2^2 + g(t) \quad (18)$$

Where λ is the Lagrangian multiplier operator and α is the quadratic penalty factor to reduce the interference of Gaussian noise. Then the alternating direction multiplier iterative algorithm and Fourier isometric transform are combined to obtain the classification and center frequency of each mode by optimization and search for the saddle point of the extended Lagrangian function, and the expressions of ω_k and λ after the alternating optimality-seeking iteration are as in Eqs. (19)-(21).

$$\hat{U}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (19)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (20)$$

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + Y [\hat{f}(\omega) - \sum_k \hat{U}_k^{n+1}(\omega)] \quad (21)$$

Where, Y is the noise tolerance to ensure that the signal is not distorted after decomposition. $\hat{u}_k^{n+1}(\omega)$, $\hat{u}_i(\omega)$, $\hat{f}(\omega)$, $\hat{\lambda}(\omega)$ are the original signal $u_k^{n+1}(t)$, $u_i(t)$, $f(t)$, $\lambda(t)$ corresponding Fourier transform. Finally, the initial signal is decomposed into K modal components according to the optimal solution.

3. Case measurement

3.1. Data sources

To verify the superiority of the CNN-BILSTM-Attention model proposed in this paper, the real energy consumption data of a middle school in Sydney, USA, and the corresponding weather data (the sampling period is 5 min, in kwh) for one month in April 2012 were selected for energy consumption prediction. The historical data from April 1 to April 24 were taken as the training set, and the data from the remaining 6 days were used as the prediction data set. At the same time, the

sequence model as shown in Figures 1 and 2 is established according to the weather time and energy consumption.

3.2. Data pre-processing

In this energy consumption prediction, there are different types of data in the dataset, such as temperature, humidity, etc., because the influence of weather data is added. To avoid the problem of different degrees of influence due to different magnitudes and to improve the accuracy of the calculation, the data need to be normalized.

In this paper, we normalize the data set to the range of [0,1] and [-1,1], and the formula is shown in equation 1).

$$y_i = (x_i - x_{\min}) / (x_{\max} - x_{\min}) \quad (22)$$

where x_i is each value in the feature vector, $x_{\min}$ is the smallest value in the feature, $x_{\max}$ is the largest value in the feature vector, and y_i is the processed data.

To examine the efficiency of the model and to assess the predictive effect, we use mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) and R-squared error (choice) as evaluation metrics. The formulas for the different evaluation metrics are represented below.

$$MPAE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (23)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (24)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (25)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (26)$$

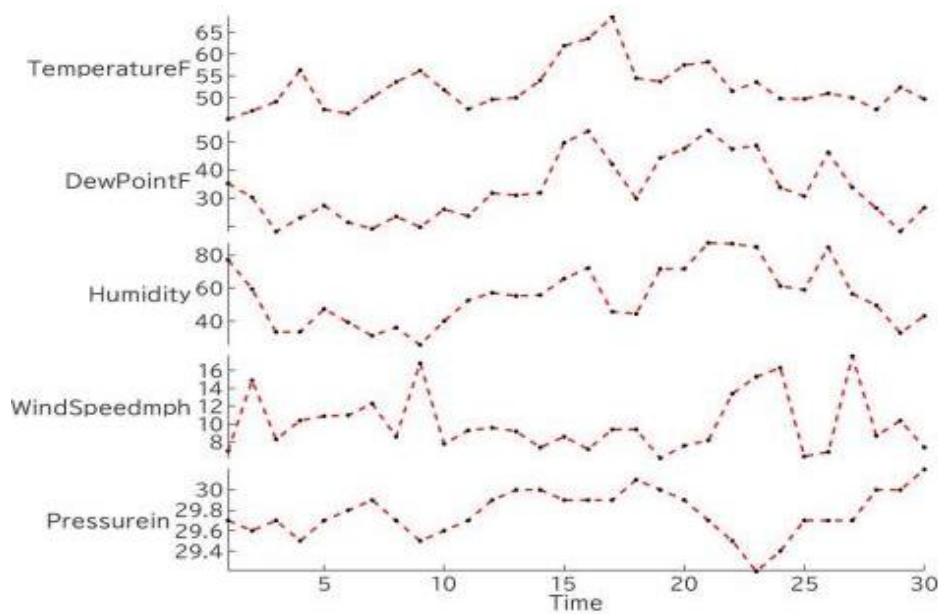


Figure 1. Weather time series model

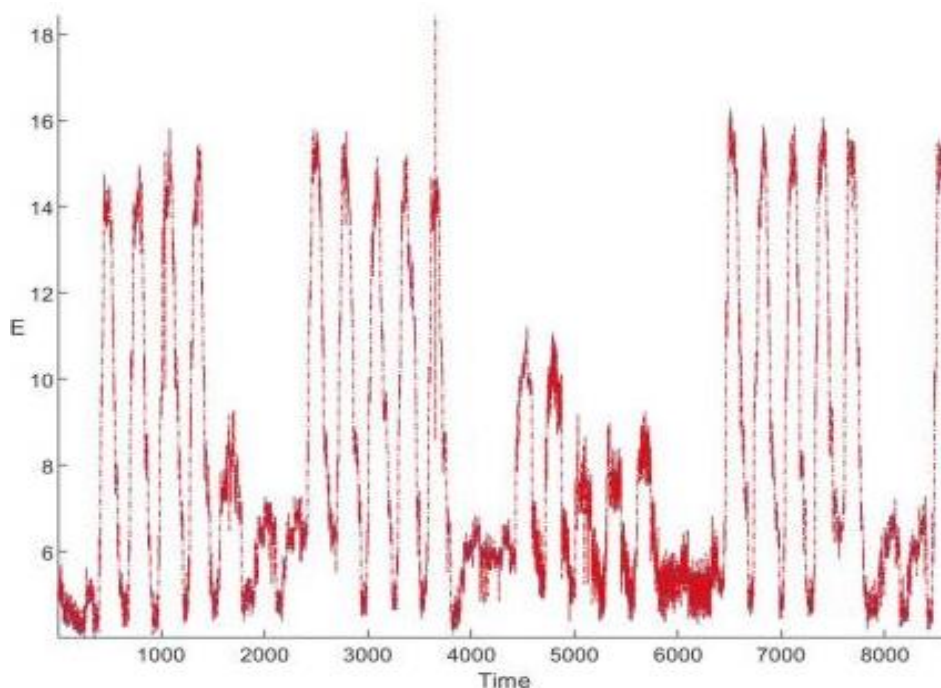


Figure 2. Energy consumption time series model

3.3. VMD Case Analysis

In this paper, VMD is applied to noise reduction and decomposition of energy consumption and weather data to improve the accuracy of the data. In this paper, the all-day building energy consumption and weather data of a region in the past month are used as the original samples for data processing. The samples of building energy consumption at different periods are smoothed using VMD to obtain several subseries with strong regularity. The sampling frequency of the processing is taken as 8640Hz, the penalty factor- α is taken as 2000, and the number of decomposition layers k is taken as 8.

As can be seen from Figure 3, the decomposition of building energy consumption using VMD yields eight components with strong regularity. IMF1 is the dominant component with a smooth curve, which can characterize the overall trend of energy loss, while the rest of the components have different frequencies but show certain regularity overall, which can reflect the local characteristics of the energy consumption sequence.

Figure 4 shows the decomposition of the VMD concerning frequency, from which it can be seen that the modal independence of each component and the larger difference in the center frequencies of the later layers indicates that it is less likely to be mixed. Figure 5 shows the variation of the original signal and the noise signal with time amplitude, from which it can be seen that the noise signal fluctuates significantly less than the original signal. In summary, it can be seen that by VMD processing, the accuracy of the data can be improved, and thus the accuracy of the model prediction can be improved.

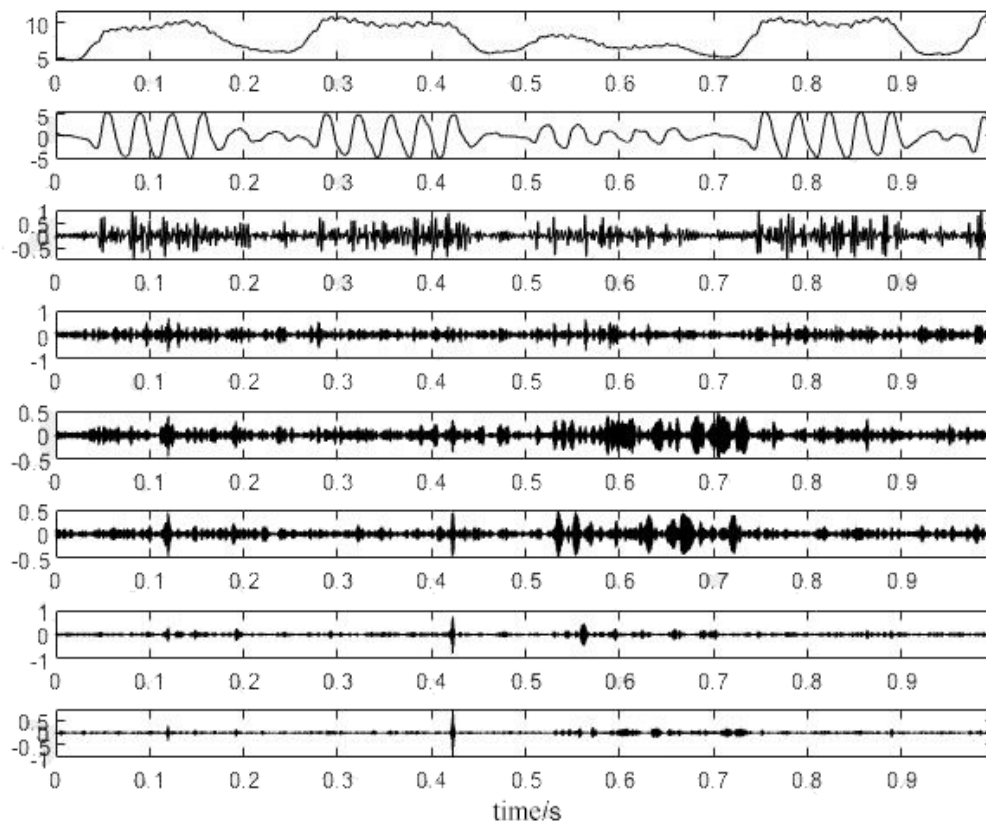


Figure 3. VMD Time exploded view

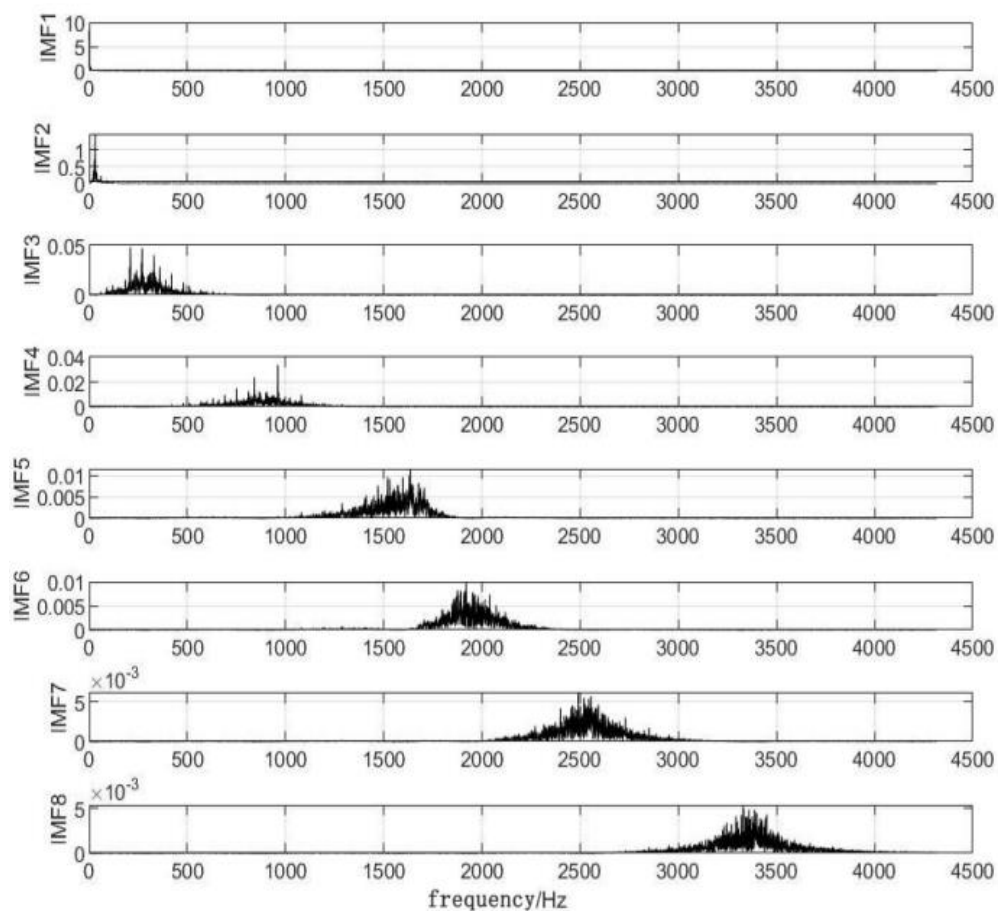


Figure 4. VMD Frequency exploded chart

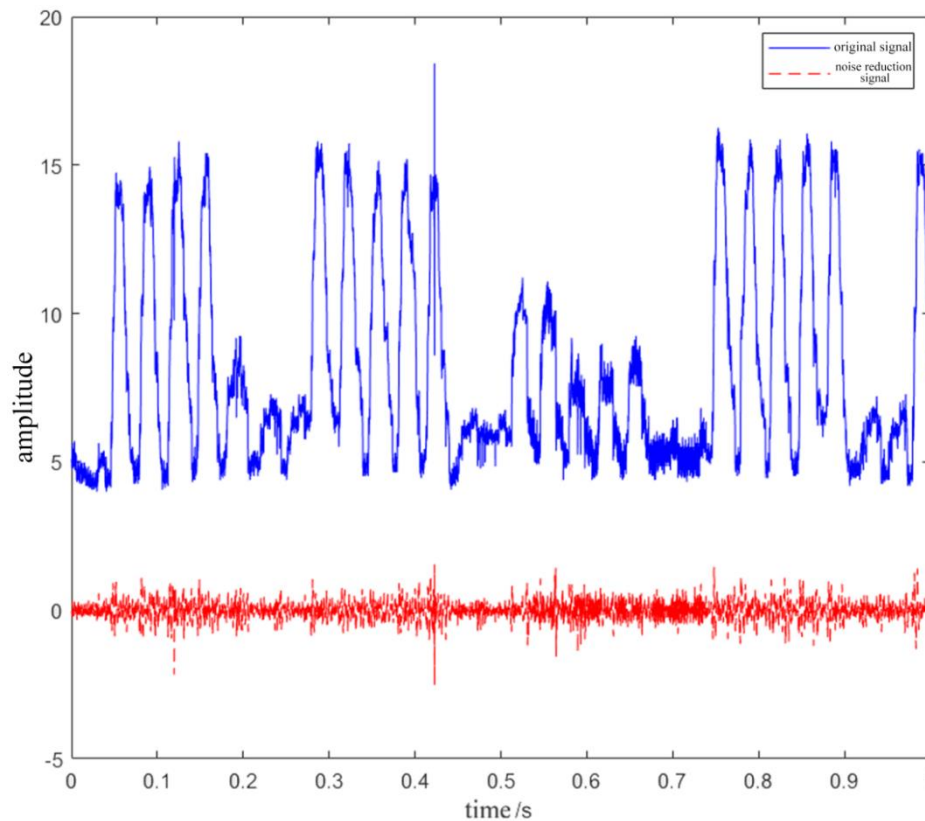


Figure 5. Original signal vs. noise signal

3.4. Analysis of energy consumption prediction results

(1) Hyperparameter settings and prediction results of CNN-BILSTM-Attention model

The proposed CNN-BILSTM-Attention model based on energy consumption prediction is implemented in Python. the CNN layer consists of a convolutional layer, a pooling layer and a fully connected layer. In the convolutional layer, the length of the convolutional kernel is set to 64 and the number of kernels in the convolutional kernel is set to 3. The pooling layer is selected for pooling with the maximum pooling principle. After extracting the feature data through the convolutional layer, it is converted into a one-dimensional vector, the BILSTM layer uses a bidirectional LSTM layer, an Attention layer is used to focus on key weather information such as temperature that affects energy consumption, and the Attention layer is multiplied with the corresponding values of the BILSTM layer and then passed through a fully connected layer. After repeated parameter tuning, 100 iterations were performed with a batch size of 256. Finally, the energy consumption prediction data is output, the result is shown in Figure 6.

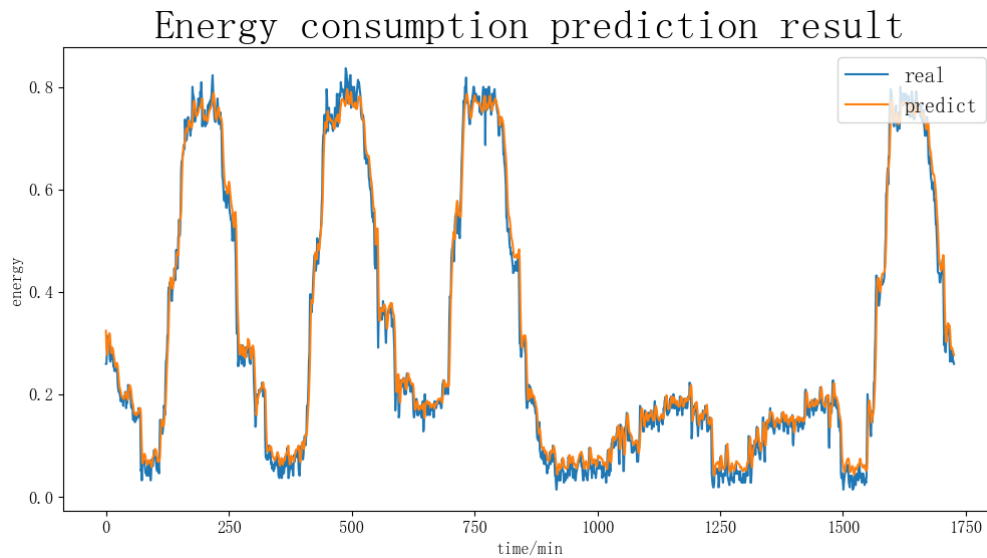


Figure 6. Energy consumption prediction results using CNN-BILSTM-Attention

(2) Comparison of prediction results by models

To confirm the efficiency and feasibility of the model, the first 80% of the normalized data were trained in five models, SVM_Regression, LSTM, CNN-LSTM, CNN-LSTM-Attention, and CNN-BILSTM-Attention, and the last 20% of the data set were used for prediction validation, and the performance of each model was compared using mean absolute error (MAE), mean absolute percentage error (MAPE), root means square error (RMSE), and R-squared error (choice) as evaluation metrics and the prediction results of each model and their corresponding metric values are shown in Figure 7.

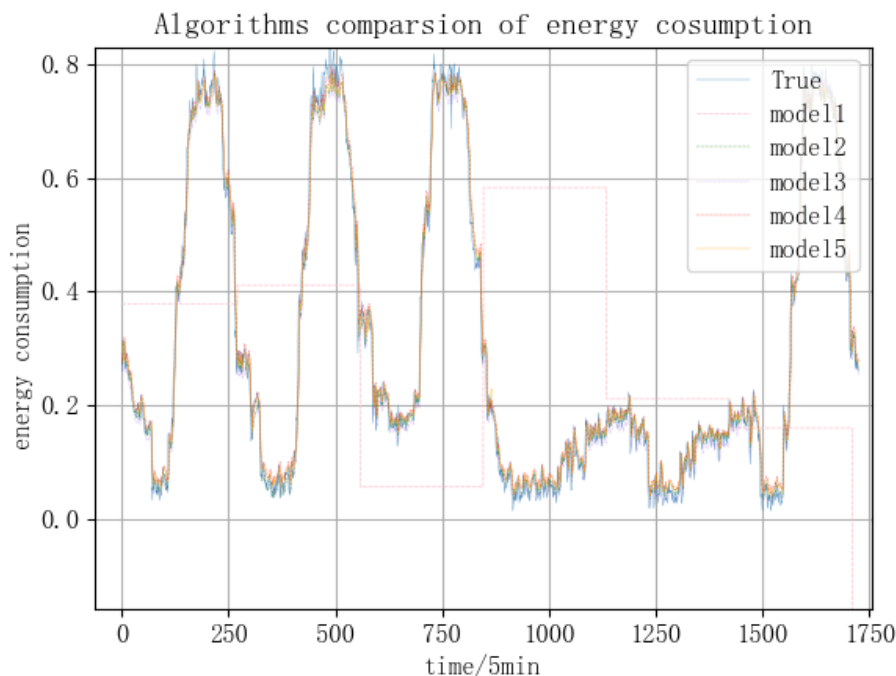


Figure 7. Result of every model

Table 1. Evaluation indicators corresponding to each algorithm

Model	Name	MAE	RMSE	MAPE	R ²
1	SVM_Regression	0.26075	0.35202	166.67	-1.1375
2	LSTM	0.02458	0.03515	14.071	0.9803
3	CNN-LSTM	0.02410	0.03387	13.734	0.9817
4	CNN-LSTM-Attention	0.02320	0.03290	13.240	0.9830
5	CNN-BILSTM-Attention	0.02212	0.03090	12.870	0.9852

1. The values of MAE and MAPE for model 1 are very large and R² negative, indicating that model 1 is a very poor fit and is not suitable for the simulation of this complex data.

2. Compared with model 1, it can be concluded that the accuracy of model 2 is much higher than that of model 1. Compared with model 1, the MAPE of model 2 decreases by 92%, and the MAE and MAPE decrease by 90.6% and 90%, respectively. The prediction results show that the traditional machine learning algorithm cannot solve the complex and large-scale data well, while the neural network model can solve this problem and reflect the dependency between individual observations.

3. Compared with model 2, the prediction accuracy of model 3 was further improved, with MAE and MAPE decreasing by 2% and 2.3%, respectively, and the data fitting accuracy R² improved by 0.27%. This is because model 3 adds a convolutional neural network CNN to model 2, which extracts effective features from the data and thus improves the accuracy.

4. Compared with model 3, the MAE and MAPE of model 4 decreased by 3.7% and 3.5%, respectively, and the RMSE also decreased by 2.8%, and the data fitting accuracy R² was further improved based on model 3. The prediction results showed that the addition of the attention mechanism could enhance the influence of important information and avoid information loss in long data series.

5. Compared with model 4, the MAE and MAPE of the fifth model CNN-BILSTM-Attention are 4.7% and 2.7%, respectively, and the RMSE is also decreased by 6%, and the data fitting accuracy R² is further improved by 0.2% based on model 3. The prediction results indicate that using BILSTM can better capture the semantic dependencies in both directions than model LSTM, bringing better fitting results and improving the prediction accuracy.

In summary, the prediction performance of different models is as follows: Model 5>model4>model3>model2>model1. The prediction model CNN-BILSTM-Attention used in this paper can effectively improve the prediction accuracy and get more accurate prediction results, and the MAPE of model 5 is reduced by 92.27% compared with the initial model 1.

4. Conclusion

With the development of China's social economy, energy saving and emission reduction are getting more and more attention. The energy consumption of buildings accounts for 25% of the total energy consumption of society, making it a major part of the national energy consumption, so energy saving in buildings is urgent. Based on this, a VMD-CNN-BILSTM-Attention neural network prediction model is proposed in this paper for building energy consumption prediction, and the following conclusions are drawn through experimental verification analysis.

(1) The VMD algorithm is first used to decompose and reduce the noise of the energy consumption data, and then normalize the data set by normalization, which improves the data accuracy and helps the subsequent neural network prediction.

(2) In the subsequent experimental data processing, the advantages of the convolutional neural network CNN, which can effectively extract spatial features, are fully utilized, combined with the BILSTM network's ability to extract bi-directional time-series features of the sequence data, and then the Attention mechanism is introduced to fully exploit the time series properties of the load data itself to distinguish the importance of different influencing factors on the energy consumption.

(3) The prediction results of four different neural network models, CNN-BiLSTM-Attention, CNN-LSTM, CNN-LSTM-Attention, and LSTM, were analyzed in the training of neural network models. The mean absolute error MAE, root mean square error RMSE, mean absolute percentage error MAPE, and coefficient of determination R^2 are used as model evaluation metrics. It was found that the CNN-BiLSTM-Attention neural network prediction model performed better in all prediction accuracy evaluation indexes and had the best prediction effect on building energy consumption.

References

- [1] A. Chel and G. Kaushik, "Renewable energy technologies for sustainable development of the energy-efficient building," *Alexandria Engineering Journal*, vol. 57, no. 2, pp. 655 – 669, Jun. 2018, doi: 10.1016/j.aej.2017.02.027.
- [2] H. Lim and Z. J. Zhai, "Review on stochastic modeling methods for building stock energy prediction," *Building Simulation*, vol. 10, no. 5. Tsinghua University Press, pp. 607–624, Oct. 01, 2017. doi: 10.1007/s12273-017-0383-y.
- [3] H. Yan, G. Ding, Yong Cheng, L. Zhang, and H. Li, "Analysis on Public Building Energy Consumption Based on Actual Data: Take 18 Buildings in Shenzhen for Example."
- [4] D. Mariano-Hernández, L. Hernández-Callejo, F. S. García, O. Duque-Perez, and A. L. Zorita-Lamadrid, "A review of energy consumption forecasting in smart buildings: Methods, input variables, forecasting horizon and metrics," *Applied Sciences (Switzerland)*, vol. 10, no. 23. MDPI AG, pp. 1–27, Dec. 01, 2020. doi: 10.3390/app10238323.
- [5] J. H. Ang, Y. Yusup, S. A. Zaki, A. Salehabadi, and M. I. Ahmad, "Comprehensive Energy Consumption of Elevator Systems Based on Hybrid Approach of Measurement and Calculation in Low-and High-Rise Buildings of Tropical Climate towards Energy Efficiency," *Sustainability (Switzerland)*, vol. 14, no. 8, Apr. 2022, doi: 10.3390/su14084779.
- [6] Divina, F., Torres, M. G., Vela, F. A. G., & Noguera, J. L. V. (2019). A comparative study of time series forecasting methods for short-term electric energy consumption prediction in smart buildings. *Energies*, 12(10). <https://doi.org/10.3390/en12101934>.
- [7] CAO Jiasheng, YANG Taihua. Application of Time Series Analysis Method in Energy Consumption Prediction of Shanghai Metro [J]. *Energy and Environment*, 2019 (3):4.
- [8] Kan Chunli. Prediction of air conditioning energy consumption based on the SARIMA GARCH model [D]. A hui Jian zhu University, 2022.
- [9] Jozi, A., Pinto, T., Praça, I., & Vale, Z. (n.d.). Day-ahead forecasting approach for energy consumption of an office building using support vector machines.
- [10] Cheng Yaohao, Chen Huanxin, WANG Jiangyu. *Refrigeration and Air Conditioning*, 2019, 19 (05): 35 - 40.)
- [11] Huang Yao. Prediction and model optimization of residential building energy consumption based on ensemble learning [D]. Huazhong University of Science and Technology, 2019. DOI: 10.27157/d.cnki.ghzku.2019.000670.
- [12] Wang, L., Deng, X., Ge, P., Dong, C., Bethel, B. J., Yang, L., & Xia, J. (2022). CNN-BiLSTM-Attention Model in Forecasting Wave Height over South-East China Seas. *Computers, Materials and Continua*, 73 (1), 2151 – 2168. <https://doi.org/10.32604/cmc.2022.027415>.
- [13] Shao, B.-L., Shi, Y.-B., & Zhao, Y. (n.d.). Research on Building Energy Consumption Prediction Model by Integrating Attention Mechanism and LSTM. <https://doi.org/10.11907/rjdk.202642>.
- [14] Chengliang Xu, Huanxin Chen, Weide Xun, Zhenxin Zhou, Tao Liu, Yuke Zeng, Tanveer Ahmad, Modal decomposition-based ensemble learning for ground source heat pump systems load forecasting, *Energy and Buildings*, Volume 194, 2019, Pages 62-74, ISSN 0378-7788, <https://doi.org/10.1016/j.enbuild.2019.04.018>.
- [15] Ruan, Y., Wang, G., Meng, H., & Qian, F. (2022). A Hybrid Model for Power Consumption Forecasting Using VMD-Based the Long Short-Term Memory Neural Network. *Frontiers in Energy Research*, 9. <https://doi.org/10.3389/fenrg.2021.772508>.

- [16] Zhao Tiecheng, nXie Lirong, Fan Xiecheng, Wang Zhiyong, Deng Yougang, Li Peng, Ye Jinxin. Prediction of polysilicon production energy consumption based on VMD and improved sparrow algorithm optimization of LSSVM [J]. Journal of Xinjiang University (Natural Science Edition), 2022, 39 (04): 498 - 507. DOI: 10.13568/j.cnki.651094.651316.2021.08.05.0001.