

District Heating System Load Prediction Using Machine Learning Method



Meng Jia, Chunhua Sun, Shanshan Cao and Chengying Qi

Abstract Accurate prediction of heating load can help improve operational efficiency of district heating systems (DHSs). The selection of feature variables is of great significance to prediction performance. Most existing methods only use the meteorological data and historical thermal demand data. In this study, correlation analysis method is employed to analyze predominant variables affecting prediction accuracy. The correlation of supply/return temperature, outdoor temperature, and historical load data were examined. The obtained results were used to select minimal input variables subset so as to avoid multiple input variables. The extreme learning machine (ELM) was used to predict the energy consumption of the next 6, 12, and 24 h. The approach was adopted to predict heating load of a DHS in Changchun, China. Historical heating load data were proved to be the most essential prediction inputs. The results show that the root-mean-square error predicted by the ELM model can reach 4.1%.

Keywords District heating system · Heating load prediction · Correlation analysis · Extreme learning machine

M. Jia · C. Sun · S. Cao (✉) · C. Qi
School of Energy and Environment, Hebei University of Technology,
Tianjin 300401, China
e-mail: css_2005@126.com

M. Jia
e-mail: hegongdajiameng@163.com

C. Sun
e-mail: sunchunhua@163.com

C. Qi
e-mail: qicy@163.com

1 Introduction

Accurate prediction of the short-term heat load is a prerequisite for efficient and stable operation of district heating system (DHS). Most existing thermal load prediction methods considered limited influencing factors, like meteorological and historical parameters, and the prediction accuracy is unstable [1–6]. These models usually reflect a smooth linear relationship between load and weather variables, which is of great nonlinearity and complexity actually [1]. A. Kusiak et al. used weather forecast data to predict steam load [2]. Nicolas Perez-Mora et al. used historical heat demand data to predict and manage DHS loads [3]. E. Dotzauer took weather forecasting and social component modeling into account [4]. H. A. Nielsen et al. obtained a regression equation between meteorological parameters (i.e., outdoor temperature, solar radiation, relative humidity, and wind speed) and building heat consumption [5]. O. Yetemen et al. found that the monsoon circulation has some influence on the long-term energy consumption prediction [6].

With the continuous development of machine learning theory, nonlinear prediction methods have been successfully applied in the field of load forecasting. Huang et al. [7] developed extreme learning machine (ELM), which is an evolutionary neural network method with good generalization ability. Sajjadi et al. established a DHS thermal load prediction model by using ELM method, revealing the robustness of this method, [8].

This paper studied the correlation of historical heating load, historical secondary supply/return temperature, and outdoor temperature. The selected input variables were used to predict heat load for the next 6, 12, and 24 h using ELM method. The proposed method was applied and analyzed in a DHS in Changchun, China.

2 Data Preprocessing

2.1 Data Outlier Elimination

Test values with coarse errors are called outliers, which are undesirable and should be removed from the measured data [9]. PauTa criterion is commonly used to judge the gross error, whose basic idea is that any error beyond triple standard deviation limit is considered to be gross error rather than random error.

When using the PauTa criterion to judge and eliminate outliers, the average value \bar{X} and residual error $V_i = X_i - \bar{X}$ of the independent measurement column $X_i (i = 1, 2, 3, \dots, n)$ should be calculated first. The standard deviation S of the measurement column is calculated. If the residual error V_d of a measured value X_d satisfies $V_d > 3S$, it is considered that X_d is an outlier needs to be rejected.

2.2 Correlation Analysis

The selection of the characteristic variables plays a crucial role in the thermal load prediction model. Through correlation analysis, the relative factors that have a great influence on load can be taken as the input factors of the prediction model to improve accuracy. In this study, the correlation coefficient method was used to analyze the correlation between two variables. r can be calculated by Eq. (1):

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \tag{1}$$

where X and Y represent the two variables. The r is between $[-1, 1]$. A positive value of r indicates a positive correlation, vice versa. The greater the absolute value of r , the stronger the correlation.

3 Prediction Methods

3.1 Extreme Learning Machine (ELM)

ELM refers to an artificial neural network model that is developed with the improvements on single-hidden layer feedforward networks (SLFNs) [10], as shown in Fig. 1.

For M arbitrary samples (x_i, t_i) , in which $x_i=[x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$. The number of single-hidden layer nodes is \tilde{N} , the standard SLFNs model with an activation function $g(x)$ is as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g_i(a_i \cdot x_j + b_i), j = 1, \dots, N \tag{2}$$

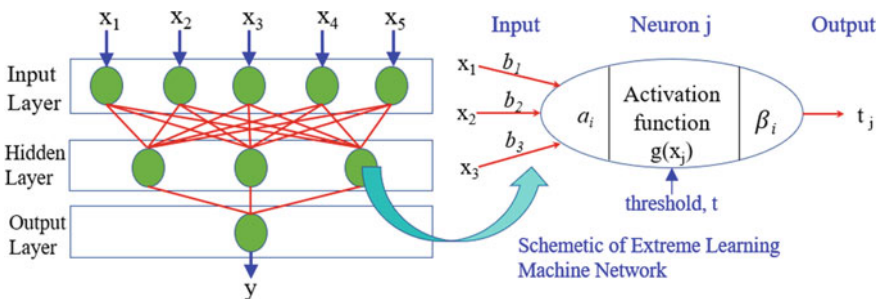


Fig. 1 Schematic of ELM network

where $a_i = [a_{i1}, a_{i2}, \dots, a_{im}]^T$ is the weight vector that connects the i th hidden layer node; b_i is the threshold of i th hidden layer nodes; $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the output weight vector connecting i th hidden layer nodes; $a_i \cdot x_j$ represents the inner product of a_i and x_j .

The ELM model can approach the output value t_j of N training samples with zero error, and we get:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(a_i \cdot x_j + b_i) = t_j, j = 1, \dots, N \tag{3}$$

Equation (4) is written in the matrix form as follows:

$$\beta H = T \tag{4}$$

where H is the hidden layer output matrix of the network; the i th column represents the output vector of the i th hidden layer node associated with the input x_1, x_2, \dots, x_N , and the j th row represents the implicit layer output vector associated with the input. The hidden layer matrix H is a deterministic matrix, so training SLFNs is equivalently converted to a least-squares solution, so that $\beta H = T$, which is expressed as follows:

$$\hat{\beta} = \min_{\beta} \left\| T(a_1, \dots, a_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}) \beta - T \right\| \tag{5}$$

Equation (6) can be expressed as follows:

$$\hat{\beta} = H^+ T \tag{6}$$

where H^+ is the molar generalized inverse matrix of the hidden layer output matrix.

3.2 Prediction Model Performance Evaluation Criteria

The mean absolute percentage error (MAPE) and root-mean-square error (RMSE) are used to evaluate the performance of the thermal load prediction model, which are relative and absolute indicators, respectively. They can be calculated by Eq. (7):

$$\begin{cases} \text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{\text{observed}_t - \text{predicted}_t}{\text{observed}_t} \right| \times 100\% \\ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\text{observed}_t - \text{predicted}_t)^2} \end{cases} \tag{7}$$

where observed_t is actual heat load and predicted_t is the predicted heat load.

4 Results and Discussion

In order to verify the feasibility and effectiveness of the proposed prediction algorithm, field test of a DHS station in Changchun City was conducted from October 21 to December 7, 2018. Outdoor temperature t_w , supply temperature t_g , return temperature t_h , and heating load q were collected every 10 min, and a total of 6840 data were collected, as shown in Fig. 2. It can be seen that t_g and t_h are relatively stable. t_w and q fluctuate more severely, which may have a certain impact on the later prediction accuracy. The measured variables were averaged every 6, 12, and 24 h, to study different timescale heat load predictions.

4.1 Correlation Analysis

The measured factors were normalized and then calculate the correlation coefficient with heat consumption according to Eq. (1), and the results are shown in Tables 1, 2, and 3.

As shown in Table 1, when the heat consumption prediction period is 6, 12, and 24 h, the historical heat consumption and the historical secondary return temperature have a strong correlation with the heating load. The correlation coefficient of historical heat consumption, historical secondary return temperature, and heating load reached the maximum when the prediction period is 12 h.

When the prediction period is 6h, 12h, 24h, the correlation coefficient between heating load and outdoor temperature is -0.485, -0.523, -0.561, respectively. Although the correlation between outdoor temperature and heating load is weak, it is the key factor in updating the heating load prediction model. Finally, we use historical heating load, secondary return temperature, and outdoor temperature as the heating load variables with prediction periods of 6, 12, and 24 h.

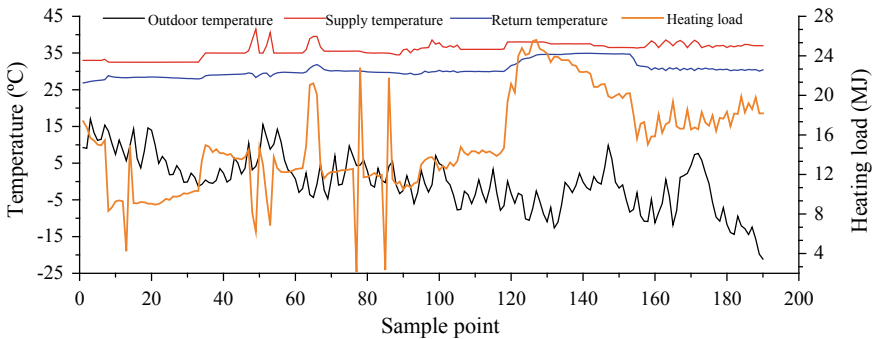


Fig. 2 Measured data

Table 1 Correlation coefficient of 6-h averaged heat load and measured data

Variable	$t_{w,i}^{(6h)}$	$q_{i-1}^{(6h)}$	$q_{i-2}^{(6h)}$	$q_{i-3}^{(6h)}$	$t_{g,i-1}^{(6h)}$	$t_{g,i-2}^{(6h)}$	$t_{g,i-3}^{(6h)}$	$t_{h,i-1}^{(6h)}$	$t_{h,i-2}^{(6h)}$	$t_{h,i-3}^{(6h)}$
$r^{(6h)}$	-0.485	0.806	0.804	0.778	0.652	0.643	0.610	0.767	0.736	0.712

Table 2 Correlation coefficient of 12-h averaged heat load and measured data

Variable	$t_{w,i}^{(12h)}$	$q_{i-1}^{(12h)}$	$q_{i-2}^{(12h)}$	$q_{i-3}^{(12h)}$	$t_{g,i-1}^{(12h)}$	$t_{g,i-2}^{(12h)}$	$t_{g,i-3}^{(12h)}$	$t_{h,i-1}^{(12h)}$	$t_{h,i-2}^{(12h)}$	$t_{h,i-3}^{(12h)}$
$r^{(12h)}$	-0.523	0.923	0.850	0.781	0.687	0.643	0.640	0.812	0.760	0.710

Table 3 Correlation coefficient of 24-h averaged heat load and measured data

Variable	$t_{w,i}^{(24h)}$	$q_{i-1}^{(24h)}$	$q_{i-2}^{(24h)}$	$q_{i-3}^{(24h)}$	$t_{g,i-1}^{(24h)}$	$t_{g,i-2}^{(24h)}$	$t_{g,i-3}^{(24h)}$	$t_{h,i-1}^{(24h)}$	$t_{h,i-2}^{(24h)}$	$t_{h,i-3}^{(24h)}$
$r^{(24h)}$	-0.561	0.875	0.759	0.613	0.664	0.640	0.638	0.777	0.687	0.611

4.2 Prediction Analysis

The data sets are divided into two categories by setting the number of test sets: the number of training sets = 7:3. As the ELM method is used to predict the heating load of the periods of 6, 12, and 24 h, the results are shown in Figs. 3, 4, and 5, respectively. It can be seen that when the predicted period of heating load is 6, 12, and 24 h, the corresponding MAPE values are 4.1, 6.8, and 9.3%. The corresponding MSE value is 0.941, 1.459, and 2.063. Comparing the prediction results, it is found that the heating load prediction model has the best degree of agreement in 6 h, the 12-h result is the second, and the 24-h fitting degree is the worst.

When the predicted period of heating load is 6 h, the trend of the predicted load curve is similar to the actual load trend. At 1–20 and 35–40 sample points, the

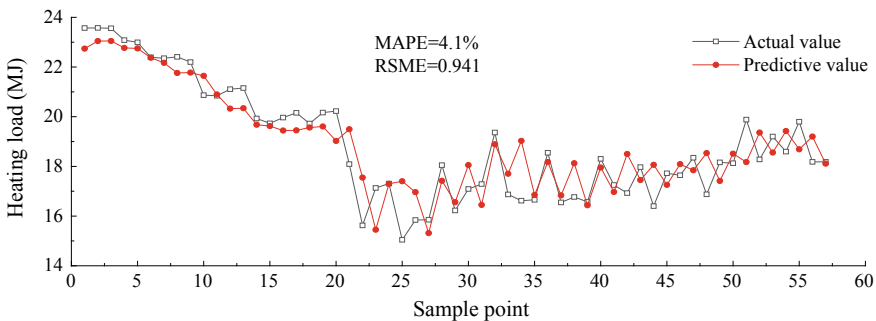


Fig. 3 Next 6-h heating load prediction results

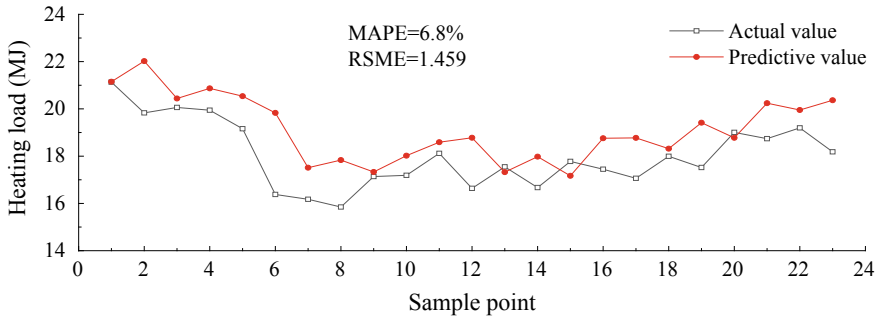


Fig. 4 Next 12-h heating load prediction results

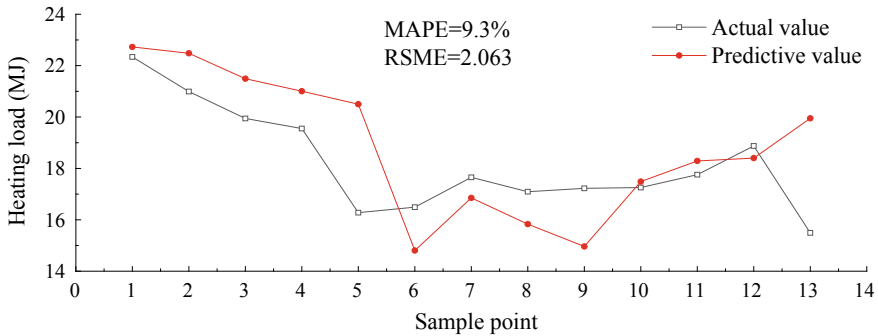


Fig. 5 Next 24-h heating load prediction results

predicted value is closer to the true value. The prediction results show that the ELM method has effectiveness in the application of short-term heating load prediction research.

With the extension of prediction time, the accuracy of heating load prediction decreases gradually. The main reason may be that the collected data samples are located in the early stage of heating, the heating load fluctuates greatly, and the collected heating load and other data are insufficient.

5 Conclusions

In this paper, the method of ELM heating load prediction is studied and verified in a heating network in Changchun. Through the establishment of ELM prediction model, the following conclusions can be drawn:

- (1) Studying the influence of different characteristic variables on heat load prediction, the MAPE values of predicted future heating loads at 6 and 12 h are 4.1 and 6.8%. It is proved that the optimized feature set model has good prediction performance.
- (2) In this study, the accuracy of the future 24-h heating load prediction is lower than the heat load forecast for the future 12 and 6 h, and its improvement measures need to be further researched.

References

1. Islam, S.M., et al.: Forecasting monthly electric load and energy for a fast growing utility using an artificial neural network. *Electr. Power Syst. Res.* **34**(1), 1–9 (1995)
2. Kusiak, A., et al.: A data-driven approach for steam load prediction in buildings. *Appl. Energy* **87**(3), 925–933 (2010)
3. Perez-Mora, N., et al.: DHC load management using demand forecast. *Energy Procedia* **91**, 557–566 (2016)
4. Dotzauer, E.: Simple model for prediction of loads in district-heating systems. *Appl. Energy* **73**(3–4), 277–284 (2002)
5. Nielsen, H.A., et al.: Modelling the heat consumption in district heating systems using a grey-box approach. *Energy Build.* **38**(1), 63–71 (2006)
6. Yetemen, O., et al.: Climatic parameters and evaluation of energy consumption of the Afyon geothermal district heating system, Afyon, Turkey. *Renew. Energy* **34**(3), 706–710 (2009)
7. Huang, G.B., et al.: Extreme learning machine: theory and applications. *Neurocomputing* **70** (1–3), 489–501 (2006)
8. Sajjadi, S., et al.: Extreme learning machine for prediction of heat load in district heating systems. *Energy Build.* **122**, 222–227 (2016)
9. Zhang, M., Yuan, H.: PauTa criteria and data outlier elimination. *J. Zhengzhou Univ. (Eng. Sci.)* **1**, 87–91 (1997)
10. Bilhan, O., et al.: The evaluation of the effect of nappe breakers on the discharge capacity of trapezoidal labyrinth weirs by ELM and SVR approaches. *Flow Meas. Instrum.* **64**, 71–82 (2018)