SHORT TERM LOAD PREDICTION OF REGIONAL HEATING AND HEAT STORAGE SYSTEM BASED ON NEURAL NETWORK

by

Yang LIU and HuiJie LIU *

Shijiazhuang Institute of Railway Technology, Shijiazhuang, Hebei, China

Original scientific paper
https://doi.org/10.2298/TSCI2402183L

Accurate heat load prediction is the key to achieve fine control, energy conservation, and carbon reduction of regional hydronics. Taking the regional hydronics of a city in the north of China as the research object, the author, respectively uses back propagation neural network (BPNN), genetic algorithm (GA) optimized BPNN (GA-BPNN), and autoregressive integrated moving average model (ARIMA) combined BPNN (ARIMA BPNN) to predict its heat load, and compares the accuracy and applicability of each prediction method. The results indicate that GA-BPNN has the smallest prediction error, followed by ARIMA-BPNN, but the latter requires less data for prediction. In practical engineering, if there is a sufficient amount of data related to heat load, it is recommended to use GA-BPNN. If there is a small amount of data, ARIMA-BP prediction method can be used.

Key words: *differential autoregressive moving average model, load prediction, BPNN, GA, district heating*

Introduction

In recent years, most regions have experienced severe haze weather. According to the measured data from national and local environmental monitoring stations, winter is a period of frequent occurrence of severe haze weather [1]. The large amount of fossil fuels consumed to meet heating needs is one of the main sources of winter pollutants. The carbon emissions from heating in the north account for 25%, while fossil fuels such as coal and natural gas account for 28%. In the sources of nitrogen oxide emissions in December, January, and February each year, Beijing and Tianjin have ranked first in emissions related to heating, surpassing emissions from transportation and industrial sectors. This is mainly due to the large amount of fossil fuels burned in winter. How to significantly reduce the emissions of pollutants in the heating industry has become an urgent issue that the government and enterprises must seriously face.

With water as the heat transfer medium, the regional hydronics, which is composed of heat source, primary pipe network, heat exchange station, secondary pipe network, indoor pipe-lines of heat users and heat dissipation ends, is the main form of heat supply in northern China. This form of hydronics conforms to the energy structure dominated by coal, which is suitable for the construction characteristics of cities, especially large and medium-sized cities, and is conducive to the realization of c ogeneration and the use of industrial waste heat for heating in the future. However, the existing regional Hydronics still faces a series of problems. The hydronics in the planning, design and actual operation process does not match, leading to difficulties in system regulation. The informatization level of hydronics is low, and the real-time

* Corresponding author, e-mail: 20150435303@mail.sdufe.edu.cn

and accuracy of basic data and operation data are low. The professional knowledge and skills of system management personnel are insufficient, relying solely on experience to *see the sky and fire*, lacking effective control indicators as a reference basis, lacking a reasonable linkage between heat source heating volume and demand side, and even artificially increasing heat source heating volume to reduce user complaints, this often leads to problems such as system thermal imbalance, excess heating, high operating energy consumption, serious resource waste, and low overall system efficiency. Therefore, in the face of increasingly tight energy resources and increasing environmental pressure today, the regional hydronics should achieve *on-demand heating* on the premise of ensuring reliable heating, improve heating quality and system energy efficiency, and reduce system energy consumption as the main research direction of regional heating practitioners. Accurate planning and control of heat load supply is an important prerequisite for ensuring heating quality and energy conservation and emission reduction, and has important research significance.

Literature review

With the rapid development of centralized heating, the energy consumption of heating accounts for an increasing proportion of the total energy consumption in society. The demand for energy is increasing day by day, but the per capita energy consumption is very limited. Therefore, heating energy conservation is one of the potential and effective ways in building energy conservation work. Winter heating not only consumes a lot of energy, but also has low energy efficiency [2]. Correct prediction and evaluation of energy consumption in heating stations can provide a basis for heating energy conservation. Heating stations are the key to meeting users' heating needs. In order to comprehensively and systematically evaluate the energy consumption level of thermal power stations, it is necessary to comprehensively consider numerous influencing factors (indicators), and multiple indicators also increase the complexity of analyzing problems to a certain extent. Domestic and foreign scholars have conducted extensive research on the influencing factors and prediction methods of heating load [3].

The time series method has a slower response to sudden weather changes and other situations. The accurate selection of independent variables in regression analysis has a significant impact on the prediction results. With the rapid development of intelligent science and technology, more and more studies have been carried out on the prediction of heat load of Hydronics using BPNN algorithm. However, BPNN method is prone to slow rate of convergence and local extremum problems. In the past, regression analysis method was often used for load forecasting of thermal power stations, but it was difficult to process data in the face of multiple factors and non-linear conditions. However, using the BPNN algorithm alone to predict heating load has low accuracy in the prediction results. Deng *et al*. [4] proposed a recurrent NN-based wireless network prediction model (BGCP-RNN-ReLU model) with Bayesian Gaussian tensor decomposition and linear correction, which can predict upstream and downstream network changes in the near future. As the demand side has developed rapidly in recent years, realistic load forecasting can better match the load of regional product and application groups. The machine learns to estimate and estimate based on the time before the forecast, which can lead to incorrect predictions of changes. Due to the rapid development of deep learning in recent years, extensive research has been conducted in the field of load prediction. Based on this, Xuan *et al*. [5] first used feature selection algorithm based on random forest, which became the basis of feature selection strategy for load estimation model. Dong *et al*. [6] proposed an improved energy integration technology including predictive models and different types of energy efficiency to improve the accuracy of load forecasting and control management of different power in mixed electricity. Using NN to generate short-term load test models and predict user needs.

From the aforementioned literature review, it can be seen that most scholars currently focus on studying the predictive effectiveness of heat load prediction methods, but there is little research on the adaptability of prediction methods to data. Under actual conditions, some data may be missing or abnormal, and the number and type of samples will change. Therefore, conducting research on the impact of changes in sample number and type on the accuracy of prediction results is of great significance. Based on the load related data of a regional hydronics in northern China, the author uses standard BPNN, GA-BPNN, and ARIMA-BPNN models to carry out load forecasting research, and compares and analyzes the applicable scenarios of each forecasting method.

Load forecasting methods and processes

The back propagation neural network

A BPNN is a multi-layer feedforward network learned based on error BP. It adjusts the connection weights by correcting the error between the generated target and the actual error to achieve the correct prediction of the network. Figure 1 shows the basic structure of the BPNN, x_j is the input value of the jth node in the input layer, θ_i the i^{th} node threshold of the hidden layer, O_k is the output value of the k^{th} node in the output layer, and w_{ij}, w_k are the interlayer weight value.

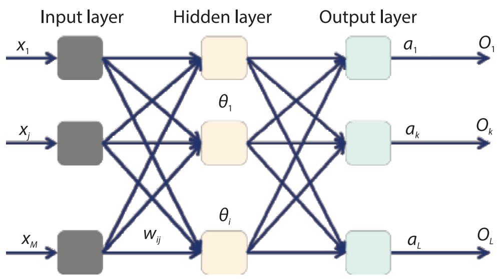

Figure 1. Schematic representation of the BPNN structure; (a) input layer, (b) hidden layer, and (c) output layer

– *The GA-BP prediction method*: The initial weights and thresholds of the standard BPNN are randomly given, which greatly affects the final prediction results. The GA is a heuristic algorithm used to solve global search optimization problems, encoding optimization variables into chromosomes and generating chromosome information through operations such as selection, crossover, and mutation, ultimately producing chromosomes that meet the optimization objectives. Firstly, the optimal weights and thresholds of the BPNN are calculated using GA, and then load forecasting is carried out:
– *The ARIMA-BP prediction method*: The ARIMA BP prediction method flowchart is shown in fig. 2. Firstly, the ARIMA model is used to predict the original heat load data, obtain the prediction results, calculate the prediction residual, analyze the linear sequence of the orig-

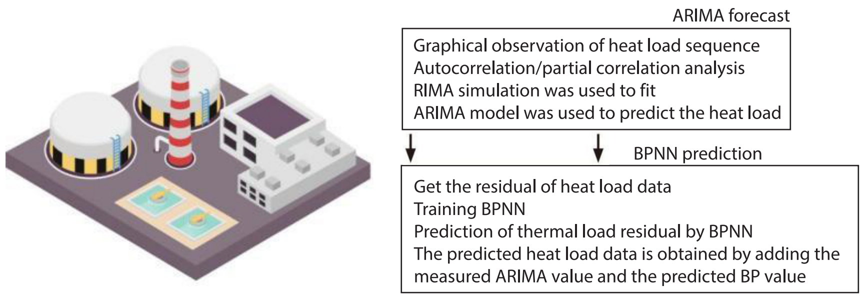

Figure 2. Flowchart of the ARIMA-BP prediction method

inal heat load data, use the BPNN prediction model for residual prediction, and deconstruct the non-linear sequence of the original load data. Add linear and non-linear sequences to obtain the final prediction result.

Divide the information contained in heat load data into linear change information and non-linear change information, and use the ARMIA model to predict the linear change sequence of heat load data. Use the BPNN prediction method to predict the non-linear sequence. The basic principle:

$$Y_t = R_t + B_t \tag{1}$$

where Y_t is the comprehensive heat load prediction sequence, R_t – the linear sequence predicted by the time series prediction model, and B_t – the non-linear sequence predicted by the BPNN prediction model [7].

– *Correlation analysis*: In order to achieve accurate prediction of heat load, the specific factors that affect heat load should first be determined. The main method is to quantitatively analyze the factors that affect heat load using correlation analysis, and the correlation coefficient calculation is shown in fig. 2. When two sets of data cannot be determined to be strictly correlated or uncorrelated, it is necessary to conduct a significance level analysis. The significance level analysis of correlation is mostly based on the set critical values, namely confidence intervals. The commonly used confidence intervals are 99%, 97.5%, and 95%, representing the significance level under this probability:

$$r_{x,y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} \tag{2}$$

$$r_{x,y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{3}$$

where $r_{x,y}$ is the correlation coefficient between variables x and y, cov(x, y) – the variable covariance, σ_x – the standard deviation of variable x, and σ_y – the standard deviation of variable y.

– *Evaluation indicators*: The prediction effect is evaluated using absolute error, relative error, and root mean square relative error:

$$E_a = |y_i - t_i| \tag{4}$$

$$E_r = \frac{|y_i - t_i|}{y_i} \times 100\% \tag{5}$$

$$RMSER = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - t_i}{y_i}\right)^2} \tag{6}$$

where E_r is the relative error and $RMSER$ – the root mean square relative error.

Case analysis

– *Data source*: A sample data of 28 days from January 4th to January 31st of a certain year in a heating area of a northern city in China was selected for research and analysis. Among them, 22 days of sample data from January 4th to January 25th were used for training, and six days of sample data from January 26th to January 31st were used for prediction [8].

– *Correlation analysis results*: Elements that may affect the temperature such as daily average outdoor wind speed, daily average outdoor temperature, daylight hours, date type, temperature of the last three days, heat load of the last two days, and heat of the previous day. The 95 check the interaction with daytime heat by looking for the reliability of *T*-values. The results of the correlation analysis are shown in tab. 1. The correlation between the average daily outdoor air temperature and wind speed is close to –1 and 1, indicating a significant effect. The correlation coefficient between solar time and the last three days' load is small, but the *T* statistic is greater than the *T*-value, so it has a significant effect. The correlation coefficient of the date mode is –0.1658, which is weak and the *T* test is in the range of positive and negative *T*-values. Therefore, the relationship between date and heat load is not significant.

Table 1. Load correlation analysis

Factor	Correlation coefficient	*T*-statistic	Relativity
Outdoor daily average temperature room	–0.9778	–3.5407	Negative linear correlation
Average wind speed outside the day	0.9601	24.4716	Positive linear correlation
Sunshine duration	–0.7850	–2.9860	Significant negative correlation
Date type	–0.1658	–0.7872	No significant correlation
Heat load for the first three days	0.4210	2.9141	Significant positive correlation
Heat load for the first two days	0.4243	2.8326	Significant positive correlation
Heat load of the previous day	0.5678	4.4036	Significant positive correlation

– *Model parameter settings*: The results of the correlation analysis include daily average outdoor wind speed, daily average outdoor temperature, time of day, date type, last three days' heat transfer, previous two days' heat transfer, and previous day's heat transfer. selected as input parameters. In the traditional BPNN algorithm, the NN structure is defined as 7-15-1 based on the calculation formula of 2m1 hidden nodes. The maximum number of training steps during training is set to 2000, the training accuracy is 0.01, and the training speed is 0.1. In the GA parameter analysis, the population size was set to 50, the genetic algebra was set to 100, the crossover probability was set to 0.95, and the variation factor result was set to 0.09. The parameters of the ARIMA algorithm should be regularly updated based on the data, and finally, the number of autoregression points, the number of variables, and the number of moving averages should be determined [9].

Result analysis

Heat load prediction based on neural network

To speed up the BPNN, data is always available before training. During training, data is divided into training set, validation set, and test set. The accuracy of the training results is represented by the regression coefficient, *R*. The closer the *R* value is to 1, the better the training. During training, the regression coefficient increased to 0.8934, indicating a good overall regression effect. A trained BPNN was used to predict heat load between January 26 and January 31.The prediction results that the BPNN predicted model is similar to the true thermal model and the prediction results are good.

Load of ARIMA-BPNN

Firstly, time series modelling was conducted for the heat load series. After multiple combination attempts, the number of autoregressive terms, *p*, was determined to be 3, the number of differences, d, was determined to be 1, and the number of moving average terms, q, was determined to be 3, indicating that the model was determined as ARIMA (3, 1, 3), the fitting of the heat load from January 4th to January 25th using this model. From the statistical fitting data of the model, the accuracy of ARIMA (3, 1, 3) prediction still needs to be improved.

The model was used to predict the data from January 26th to January 31st, with a confidence level of 95%. The prediction effect of the model is shown in fig. 3. From the figure, it can be seen that, the overall heat load prediction results obtained from the ARIMA prediction model are relatively high and have a flat trend. This result is a linear series of heat load time series, but it is far from the actual heat load values. Therefore, it is necessary to further deconstruct the non-linear series of heat load series.

Extract the heat load data obtained by ARIMA model installation, predict the actual value of heat load, obtain the waste heat data, and build the BPNN model. Predict daily average outdoor temperature, daily average outdoor wind speed, heat load sequence at high confidence level, and thermal sequence with low confidence according to four measures of BPNN strategy. The number of nodes in the hidden layer can be determined by the principle of 2m1, and the output layer is the residual heat load, so the 4-9-1 model of the BPNN is created. After training, predict the data from January 26 to January 31 as shown in fig. 4. As shown in the figure, the prediction results of other heats using BPNN are similar to reality: heat source model, low overall prediction error, and prediction accuracy.

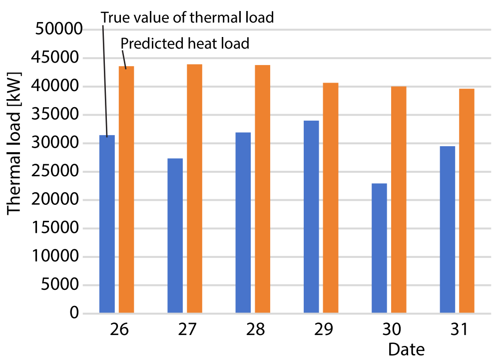

Figure 3. Prediction effect of ARIMA model

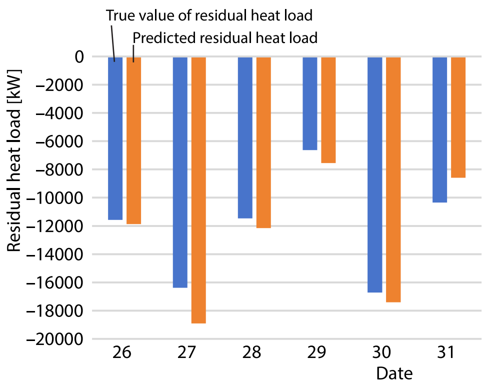

Figure 4. Results plot of heat load residual prediction

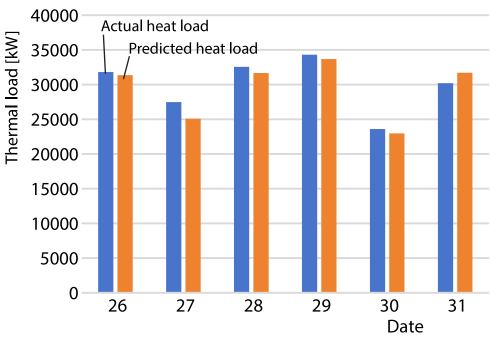

Figure 5. The ARIMA-BP heat load prediction results

By combining the linear part of the ARIMA predicted heat load sequence with the non-linear part of the BPNN predicted heat load sequence, the ARIMA-BP prediction results are obtained, as shown in fig. 5. From the figure, it can be seen that after adding the linear part of the ARIMA predicted heat load sequence to the non-linear part of the BPNN predicted heat load sequence, the predicted results of the ARIMA-BPNN are similar to the true trend of heat load, and the overall prediction effect is good.

Comparison of prediction errors

Compare the predictive performance of BPNN, GA-BPNN, ARIMA, and ARIMA-BPNN. The relative errors between the predicted values and actual values of each model are shown in fig. 6. As shown in the figure, ARIMA has the worst accuracy in predicting heat load, followed by BPNN, and GA-BPNN has the most accurate prediction [10]. Table 2 shows the error statistics between the predicted and actual values of the model. It can be seen that the average relative error and RMSER of ARIMA-BPNN are only 0.78% and 0.95% larger than GA-BPNN. However, ARIMA-BPNN only uses three types of data: Heat load, outdoor daily average temperature, and outdoor daily average wind speed, while GA-BPNN uses seven types of data. Therefore, in practical engineering, if there is enough heat load related data, it is recommended to use GA-BPNN. If there is less heat load related data, ARIMA-BP prediction method can be tried.

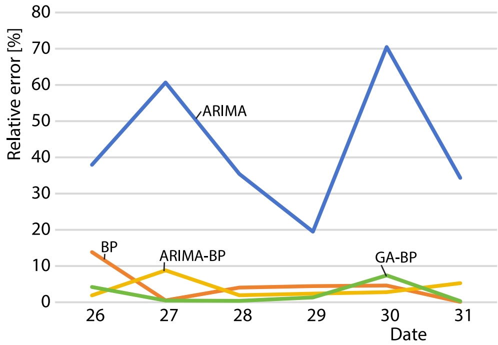

Figure 6. Comparison diagram of the prediction effect

Table 2. Comparison table of the prediction errors

Error	ARIMA	BP	ARIMA-BP	GA-BP
E_a [kW]	12344.36	1365.34	1050.73	544.19
E_r [%]	43.33	4.36	3.58	2.02
RMSER [%]	46.77	6.29	4.47	3.52

Conclusion

Proper and sound Hydronics regulation is key to saving energy and reducing carbon emissions in the heating industry, and the primary goal of implementing these regulations is to ensure Hydronics heat. The author uses BPNN, GA-BPNN method, and ARIMA-BPNN algorithm to predict hydronic zone heat of a city in northern China. The main conclusion is GA-BPNN and ARIMA-BPNN are better than neural network prediction method B, among them, the average relative prediction error of BPNN is 4.36 and GA-BPNN is 4.36 and ARIMA-BP is 3.58 Compared to GA-BPNN, ARIMA-BPNN requires less data. For technology, GA-BPNN is recommended if there is enough data related to heat load. If the data is sparse, ARIMA-BP forecasting method can be used.

References

[1] Wang, Y., *et al.*, Short-Term Load Prediction of Integrated Energy System Based on Neural Network, *Journal of Physics: Conference Series*, *2087* (2021), 1, 012016

[2] Zixu, G., Yi C., Thermal Load Analysis and Control of Four-Stroke High Speed Diesel Engine, *Thermal Science*, *25* (2021), 4B, pp. 2871-2879

[3] Li, D., Prediction of Short-Term Load of Microgrid Based on Multivariable and Multistep Long Short-Term Memory, Sensors and Materials, *An International Journal on Sensor Technology*, *85* (2022), 4, Part 1, 34

[4] Deng, T., *et al.*, Short Term Prediction of Wireless Traffic Based on Tensor Decomposition and Recurrent, *Neural Network*, *15* (2021), 4, 8547

[5] Xuan, Y., *et al.*, Multi-Model Fusion Short-Term Load Forecasting Based on Random Forest Feature Selection and Hybrid Neural Network, *IEEE Access*, *PP* (2021), 99, pp. 1-1

[6] Dong, H., *et al*., Optimized Operation of Integrated Energy Microgrid with Energy Storage Based on Short-Term Load Forecasting. Electronics, *11* (2021), 7, pp. 45-49
[7] Gao, W., *et al.*, Short-Term Cooling Load Prediction for Office Buildings Based on Feature Selection Scheme and Stacking Ensemble Model. Engineering Computations, *International Journal For Computer-Aided Engineering and Software*, *69* (2022), 5, 39
[8] Dong, L., *et al.*, Investigation of the Explosion Characteristics of Ethylene-Air Premixed Gas in Flameproof Enclosures by Using Numerical Simulations, *Thermal Science*, *27* (2023), 2B, pp. 1573-1586
[9] Guo, B., Friction Heat Energy Recovery System Based on Hydraulic Brake System by Wire of Heavy Vehicle, *Thermal Science*, *27* (2023), 2A, pp. 1159-1166
[10] Gong, M., *et al*., Load Forecasting of District Heating System Based on Informer, *Energy*, *253* (2022), 4, 1025

Paper submitted: May 2, 2023
Paper revised: July 28, 2023
Paper accepted: September 1, 2023