Energy Conversion and Management 50 (2009) 319–327

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Prediction of energy demands using neural network with model identification by global optimization

Ryohei Yokoyama *, Tetsuya Wakui, Ryoichi Satake

Department of Mechanical Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan

ARTICLE INFO

Article history:
Received 19 February 2008
Accepted 12 September 2008

Keywords:
Prediction
Energy demands
Multi-layered neural network
System identification
Non-linear programming
Global optimization

ABSTRACT

To operate energy supply plants properly from the viewpoints of stable energy supply, and energy and cost savings, it is important to predict energy demands accurately as basic conditions. Several methods of predicting energy demands have been proposed, and one of them is to use neural networks. Although local optimization methods such as gradient ones have conventionally been adopted in the back propagation procedure to identify the values of model parameters, they have the significant drawback that they can derive only local optimal solutions. In this paper, a global optimization method called "Modal Trimming Method" proposed for non-linear programming problems is adopted to identify the values of model parameters. In addition, the trend and periodic change are first removed from time series data on energy demand, and the converted data is used as the main input to a neural network. Furthermore, predicted values of air temperature and relative humidity are considered as additional inputs to the neural network, and their effect on the prediction of energy demand is investigated. This approach is applied to the prediction of the cooling demand in a building used for a bench mark test of a variety of prediction methods, and its validity and effectiveness are clarified.

1. Introduction

Distributed energy supply plants have conventionally been operated for stable energy supply based on experiences and intuitions of operators. In addition to stable energy supply, however, energy and cost savings have recently been required to operators. Therefore, it has become difficult for operators to operate the plants properly in consideration of energy demands which change with season and time as well as corresponding energy consumptions and costs.

Under these situations, advisory systems which assist operators in operating energy supply plants in real-time manner are required [1]. Main functions for the advisory systems are to predict energy demands accurately as basic conditions for the operational planning, and to conduct the operational planning properly based on the predicted energy demands from the viewpoints of stable energy supply, and energy and cost savings.

In this paper, only the former function or the prediction of energy demands is discussed. Several methods of predicting energy demands have been proposed, and typical examples are time series analysis, Grey model, and neural network [2]. Among them, the neural network and extended methods have been proposed increasingly [3–11]. In order to predict energy demands accurately using the neural networks, it is important to identify the values of model parameters. In many cases, local optimization methods such as gradient ones have conventionally been adopted in the back propagation procedure to minimize the summations of squared errors as the objective functions for this purpose. However, the summations of squared errors are generally non-convex or multimodal with respect to model parameters, and therefore, local optimization methods have the significant drawback that they can derive only local optimal solutions. To remove this drawback, the genetic algorithm and heuristic methods have been adopted, but very limitedly [8,11], and global optimization methods have not been established to identify the values of model parameters for the neural networks. In addition, time series data on energy demands have conventionally been used directly as the inputs to the neural networks in many cases. However, there is a possibility of enhancing the accuracy of the prediction by processing time series data on energy demands in advance of using the neural networks. Furthermore, it may be possible to enhance the accuracy of the prediction by considering supplementary conditions as additional inputs to the neural networks.

In this paper, the following approach is proposed to predict energy demands accurately using neural networks. The trend and periodic change are first removed from time series data on energy demand, and the converted data is used as the main input to a neural network. Here, a global optimization method called "Modal Trimming Method" proposed for non-linear programming problems is adopted to identify the values of model parameters [12]. In addition, predicted values of air temperature and relative humidity are considered as additional inputs to the neural

* Corresponding author. Tel.: +81 72 254 9229; fax: +81 72 254 9904.
E-mail address: yokoyama@me.osakafu-u.ac.jp (R. Yokoyama).

0196-8904/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enconman.2008.09.017

Nomenclature

$\hat{a}$	error between measured and predicted values of w or zero (kWh/h)
C	binomial coefficient
D	order of periodic difference
d	order of ordinary difference
F	function composed of objective function
f	objective function
g	function for neurons of hidden layer
h	function for neurons of output layer
I	number of neurons of input layer
J	number of neurons of hidden layer
K	number of neurons of output layer
L	number of sampling times for prediction
M	number of sampling times for measured energy demand
N	number of patterns
p	number of sampling times for values of $\hat{w}$ for prediction
q	number of sampling times for values of $\hat{a}$ for prediction
R	number of sampling times reduced for predicted values of w
s	number of sampling times for periodic differential operation
Δt	sampling time interval (h)
u	weights for neurons between input and hidden layers
v	weights for neurons between hidden and output layers
w	time series data after Dth order periodic differential operation on y (kWh/h)
$\hat{w}$	measured or predicted value of w (kWh/h)
X	input to and output from neurons of input layer
$\boldsymbol{x}$	vector for variables
Y	output from neurons of hidden layers
y	time series data after dth order ordinary differential operation on z (kWh/h)
Z	output from neurons of output layer
z	time series data on energy demand (kWh/h)
α	constant coefficient for h
β	step width
γ	decelerating parameter
η	thresholds for neurons of hidden layer
θ	thresholds for neurons of output layer
∇	ordinary differential operator
∇_s	periodic differential operator
$\tilde{()}_{t'\|t}$	value at sampling time t' predicted at sampling time t
$\\| \\|_2$	Euclidean norm

Subscripts

i	index for neurons of input layer
j	index for neurons of hidden layer
k	index for neurons of output layer
(m)	number of renewal of values of variables
n	index for patterns
t, l	index for sampling times

Superscripts

T	transposition
+	Moore–Penrose generalized inverse
-1	inverse
*	tentative global quasi-optimal solution

network, and its effect on the prediction of energy demand is investigated. For this purpose, the following two-phase method is tried: At the first phase, the air temperature and relative humidity are predicted independently using their measured or predicted values; At the second phase, the energy demand is predicted using its measured or predicted values as well as the predicted values of air temperature and relative humidity. This approach is applied to the prediction of the cooling demand in a building used for a bench mark test of a variety of prediction methods. The results obtained by the proposed approach are compared with those by the conventional one.

2. Prediction model

Fig. 1 shows the flow of data processing of the prediction model proposed in this paper. First, the trend and periodic change are removed from time series data on energy demand. Second, the converted data is used as the input to a neural network. Finally, the output from the neural network is converted to predicted values of energy demand.

2.1. Removal of trend and periodic change

Before the application of the neural network, the trend and periodic change are removed from time series data on energy demand. Especially, since the energy demand changes periodically with a period of 24 h, the removal of the periodic change will enhance the accuracy of the prediction. Here, the time series data on energy demand is designated by z_t ($t = 1, 2, \ldots$) with a sampling time interval of Δt.

First, the ordinary differential operation is conducted to remove the trend, and the first order ordinary difference is obtained as follows:

$$\nabla z_t = z_t - z_{t-1} \tag{1}$$

where ∇ is the ordinary differential operator. By conducting this operation d times, the dth order ordinary difference y_t is obtained as follows:

$$y_t = \nabla^d z_t = \begin{cases} z_t & (d = 0) \\ \sum_{e=0}^{d} (-1)^e {}_d\mathrm{C}_e z_{t-e} & (d \geqslant 1) \end{cases} \tag{2}$$

where ${}_d\mathrm{C}_e$ is the binomial coefficient which expresses the number of combinations for selecting e elements among d ones.

Second, the periodic differential operation is conducted to remove the periodic change, and the first order periodic difference is obtained as follows:

$$\nabla_s y_t = y_t - y_{t-s} \tag{3}$$

where ∇_s is the periodic differential operator with a period of $s\Delta t$, and s is the number of sampling times for the periodic differential operation. By conducting this operation D times, the Dth order periodic difference w_t is obtained as follows:

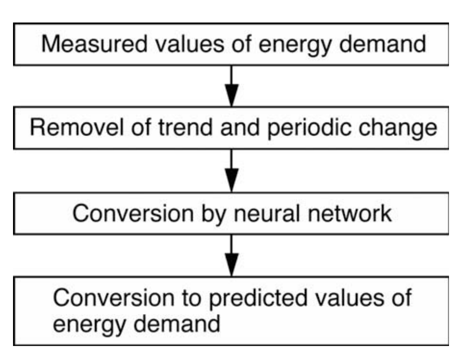

Fig. 1. Flow of data processing.

$$w_t = \nabla_s^D y_t = \begin{cases} y_t & (D = 0) \\ \sum_{e=0}^{D} (-1)^e {}_D C_e y_{t-es} & (D \geqslant 1) \end{cases} \tag{4}$$

2.2. Three-layered neural network model

As shown in Fig. 2, a three-layered neural network model is considered. This neural network model is composed of I, J, and K neurons of the input, hidden, and output layers, respectively. To apply this model to the prediction of energy demand, the number of neurons of the output layer is set at $K = 1$.

2.2.1. Input and output

As the output from the model, the value $\tilde{w}_{t+l|t}$ at the sampling times $t + l(l = 1, 2, \cdots, L)$ predicted at the sampling time t is adopted, where $\tilde{()}_{t'|t}$ means the value at the sampling time t' predicted at the sampling time t. Here, it is assumed that $L \leqslant s$. As the input to the model, the p measured or predicted values $\hat{w}_{t+l-1}, \hat{w}_{t+l-2}, \cdots, \hat{w}_{t+l-p}$ at the sampling times $t + l - 1$, $t + l - 2$, $\ldots$, $t + l - p$, respectively, are adopted. Here, $\hat{w}_{t'}$ is evaluated as the measured value $w_{t'}$ if it can be evaluated, otherwise as the predicted value $\tilde{w}_{t'|t}$. Here, it is assumed that $p < s$. In addition, the q errors between the measured and predicted values $\hat{a}_{t+l-1}, \hat{a}_{t+l-2}, \ldots, \hat{a}_{t+l-q}$ at the sampling times $t + l - 1$, $t + l - 2$, $\ldots$, $t + l - q$, respectively, are adopted. Here, $\hat{a}_{t'}$ is evaluated as

$$\hat{a}_{t'} = w_{t'} - \tilde{w}_{t'|t'-1} \tag{5}$$

if $w_{t'}$ can be evaluated, otherwise as zero. Here, it is assumed that $q < s$. This is similar to the autoregressive moving average model for time series analysis. As a result, the number of the neurons of the input layer is $I = p + q$.

2.2.2. Input–output relationships

The input–output relationships for the jth neuron of the hidden layer and the kth neuron of the output layer are expressed as follows:

$$Y_{jn} = g\left(\sum_{i=1}^{I} u_{ji} X_{in} + \eta_j\right) \quad (n = 1, 2, \ldots, N;\ j = 1, 2, \ldots, J) \tag{6}$$

$$Z_{kn} = h\left(\sum_{J=1}^{J} v_{kj} Y_{jn} + \theta_k\right) \quad (n = 1, 2, \ldots, N;\ k = 1, 2, \ldots, K) \tag{7}$$

respectively. Here, X_{in} is the input to and output from the ith neuron of the input layer for the nth pattern, Y_{jn} is the output from the jth neuron of the hidden layer for the nth pattern, Z_{kn} is the output from the kth neuron of the output layer for the nth pattern, u_{ji} is the weight for the connection between the ith neuron of the input layer and the jth neuron of the hidden layer, η_j is the threshold for the jth neuron of the hidden layer, v_{kj} is the weight for the connection between the jth neuron of the hidden layer and the kth neuron of the output layer, and θ_k is the threshold for the kth neuron of the output layer.

In Eqs. (6) and (7), N is the number of patterns, which is equal to the number of the predicted values $\tilde{w}_{t+l|t}$, or the number of the sets of the measured or predicted values $\hat{w}_{t+l-1}, \hat{w}_{t+l-2}, \ldots, \hat{w}_{t+l-p}$ and the errors between the measured and predicted values $\hat{a}_{t+l-1}, \hat{a}_{t+l-2}, \ldots, \hat{a}_{t+l-q}$.

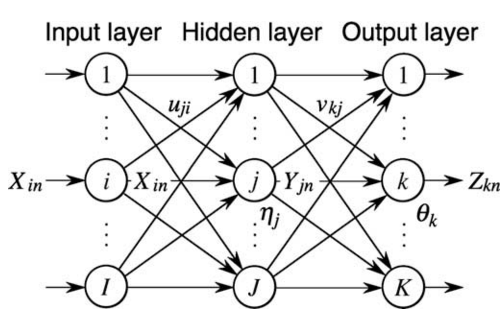

Fig. 2. Three-layered neural network model.

In Eqs. (6) and (7), $g(x)$ and $h(x)$ are the functions which convert the input into the output in the neurons of the hidden and output layers, respectively. The sigmoid function has conventionally been adopted as these conversion functions. However, the predicted value $\tilde{w}_{t+l|t}$ can be negative and its absolute value can be larger than unity in the problem under consideration. Therefore, the following functions are adopted here:

$$g(x) = \tanh x \tag{8}$$

$$h(x) = \alpha \tanh x \tag{9}$$

where α is a constant coefficient.

2.3. Conversion from output to energy demand

From Eq. (4), the relationship between the predicted values $\tilde{w}_{t+l|t}$ and $\tilde{y}_{t+l|t}$ is expressed as follows:

$$\tilde{y}_{t+l|t} = \begin{cases} \tilde{w}_{t+l|t} & (D = 0) \\ \tilde{w}_{t+l|t} - \sum_{e=1}^{D} (-1)^e {}_D C_e y_{t+l-es} & (l = 1, 2, \ldots, L)(D \geqslant 1) \end{cases} \tag{10}$$

In addition, from Eq. (2), the relationship between the predicted values $\tilde{y}_{t+l|t}$ and $\tilde{z}_{t+l|t}$ is expressed as follows:

$$\tilde{z}_{t+l|t} = \begin{cases} \tilde{y}_{t+l|t} & (d = 0) \\ \tilde{y}_{t+l|t} - \sum_{e=1}^{l-1} (-1)^e {}_d C_e \tilde{z}_{t+l-e|t} & \\ -\sum_{e=l}^{d} (-1)^e {}_d C_e z_{t+l-e} & (l = 1, 2, \ldots, L)(d \geqslant 1, l \leqslant d) \\ \tilde{y}_{t+l|t} - \sum_{e=1}^{d} (-1)^e {}_d C_e \tilde{z}_{t+l-e|t} & (d \geqslant 1, l > d) \end{cases} \tag{11}$$

The use of these equations can convert the output from the neural network model to the predicted values of energy demand.

3. Two-phase method

To enhance the accuracy of the prediction further, it may be necessary to consider supplementary conditions which affect energy demands. In this paper, the air temperature and relative humidity are considered as the most dominant supplementary conditions. There may be some methods of considering such supplementary conditions. In this paper, predicted values of air temperature and relative humidity are considered as additional inputs to the neural network model. For this purpose, a two-phase method is adopted as shown in Fig. 3. At the first phase, the air temperature and relative humidity are predicted independently using their measured or predicted values by neural networks. At the second phase, the energy demand is predicted using its measured or predicted values as well as the predicted values of air temperature and relative humidity by another neural network. Here, it is considered that the energy demand at a sampling time is affected dominantly by the air temperature and relative humidity at the same sampling time. For the prediction of air temperature and relative humidity, the number of neurons of the input layer is $I = p + q$. For the prediction of energy demand, the number of neurons of the input layer is $I = p + q + 2$.

4. Model identification

4.1. Optimization problem for back propagation

Generally, the squared error between the output from the kth neuron of the output layer Z_{kn} and the corresponding teaching data

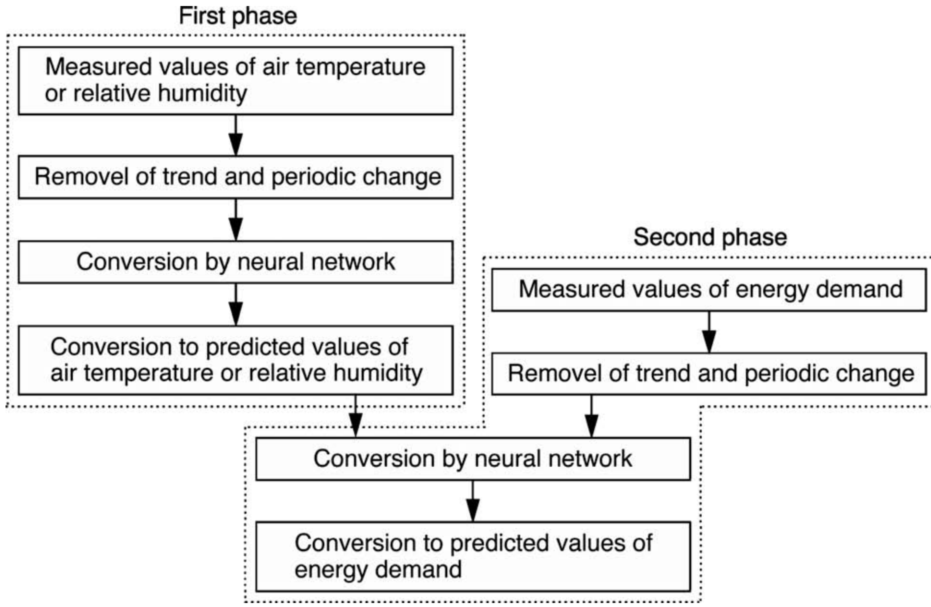

Fig. 3. Concept of two-phase method.

is evaluated for the nth pattern, and its summation for the K outputs and N patterns is minimized as the objective function to identify the values of model parameters for the neural network. In the problem under consideration, the output Z_{kn} and the corresponding teaching data are replaced by the predicted and measured values of energy demand, $\tilde{z}_{t+l|t}$ and z_{t+l}, respectively. Therefore, the objective function is defined as follows:

$$f = \sum_{l=1}^{L} \sum_{t=R}^{M-l} (\tilde{z}_{t+l|t} - z_{t+l})^2 \tag{12}$$

The number of patterns N is replaced by $L\{(M-R)-(L-1)/2\}$, where M is the number of sampling times for the measured value of energy demand, and R is the number of sampling times reduced for the predicted value $\tilde{z}_{t+l|t}$ and is expressed as follows:

$$R = Ds + d + \max(p, q) \tag{13}$$

where $\max(p, q)$ means that the larger value is selected between p and q. The variable vector composed of the weights and thresholds whose values are to be determined is defined as follows:

$$\boldsymbol{x} = (u_{11}, \ldots, u_{JI}, v_{11}, \ldots, v_{KJ}, \eta_1, \ldots, \eta_J, \theta_1, \ldots, \theta_K)^{\mathrm{T}} \tag{14}$$

where the superscript T means the transposition of a vector.

In the back propagation method, the error function for each pattern in Eq. (12) is minimized sequentially. Here, to secure the local optimality of solutions and make the convergence faster, the total error function for all the patterns of Eq. (12) is minimized simultaneously.

4.2. Search for local optimal solutions

The search for local optimal solutions can be conducted by gradient methods for unconstrained non-linear programming problems. For example, the simplest gradient method is the steepest descent one, and the renewal of the values of the variables $\boldsymbol{x}$ or the model parameters for the problem under consideration is expressed as follows:

$$\boldsymbol{x}_{(m+1)} = \boldsymbol{x}_{(m)} - \beta[\partial f(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}]^{\mathrm{T}} \tag{15}$$

where the subscript (m) is the number of the renewal of the values of the variables $\boldsymbol{x}$, and β is the step width. However, the steepest descent method has the drawback of slow convergence. Therefore, the conjugate gradient and quasi-Newton methods with faster convergence are often adopted instead of the steepest descent method.

Nevertheless, as aforementioned, these methods have the significant drawback that they can derive only local optimal solutions. Although some efforts have been made to derive global optimal solutions in identifying the values of model parameters for neural networks [13,14], global optimization methods have not been established. In this paper, the modal trimming method proposed for non-linear programming problems is adopted as a global optimization one [12]. This method has been applied to a neural network for pattern recognition, and its validity and effectiveness have been ascertained [15].

4.3. Global optimization by modal trimming method

4.3.1. Basic concept

The concept of the modal trimming method is shown in Fig. 4. This method is composed of the following two procedures: A local optimal solution is searched to obtain a tentative global quasi-optimal one; A feasible solution with the value of the objective function equal to that for the tentative global quasi-optimal one is searched to obtain an initial point for finding a better local optimal one. These procedures are repeated until a feasible solution with

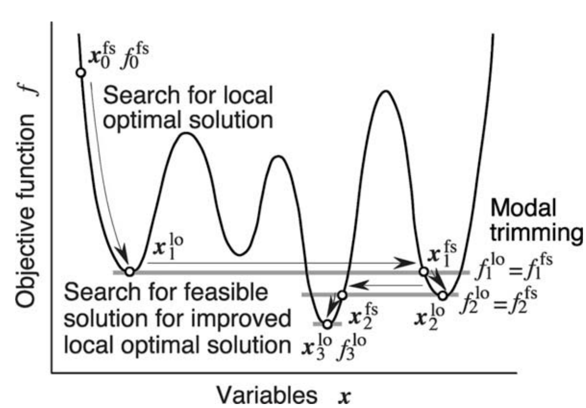

Fig. 4. Concept of modal trimming method.

the value of the objective function equal to that for the tentative global quasi-optimal one cannot be found, and the tentative global quasi-optimal one is adopted as the global quasi-optimal one.

The aforementioned idea is similar to that of the tunneling algorithm, which has also been proposed to derive global quasi-optimal solutions [16]. In the tunneling algorithm, a feasible solution is searched by a conventional gradient method. However, in the modal trimming method, a feasible solution is searched by an extended Newton–Raphson method based on the Moore–Penrose generalized inverse of the Jacobi matrix of the objective function and constraints. The method can have a high possibility of deriving global optimal solutions, if it has the capability of global search for feasible ones.

4.3.2. Search for feasible solutions

An extended Newton–Raphson method is applied to search a feasible solution with the value of the objective function equal to that for the tentative global quasi-optimal one. Namely, the function composed of the objective function

$$F(\boldsymbol{x}) = f(\boldsymbol{x}) - f^* \tag{16}$$

is defined, and the values of the variables $\boldsymbol{x}$ are renewed to satisfy $F(\boldsymbol{x}) = 0$ by the following equation:

$$\boldsymbol{x}_{(m+1)} = \boldsymbol{x}_{(m)} - \gamma[\partial F(\boldsymbol{x}_{(m)}/\partial \boldsymbol{x}]^+ F(\boldsymbol{x}_{(m)}) \tag{17}$$

In Eq. (16), f^* is the value of the objective function for the tentative global quasi-optimal solution. In Eq. (17), $[\,]^+$ means the Moore–Penrose generalized inverse and is applied to the Jacobi matrix $\partial F(\boldsymbol{x})/\partial \boldsymbol{x}$ of the function $F(\boldsymbol{x})$, and γ is the decelerating parameter whose value is between 0 and 1.

The application of the Moore–Penrose generalized inverse is because the multiple variables $\boldsymbol{x}$ produce numerous solutions which satisfy $F(\boldsymbol{x}) = 0$. For the problem under consideration, the Moore–Penrose generalized inverse is reduced to the right inverse as follows:

$$\begin{aligned}[\partial F(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}]^+ &= [\partial F(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}]^T([\partial F(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}][\partial F(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}]^T)^{-1} \\ &= [\partial f(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}]^{\mathrm{T}}/\|\partial f(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}\|_2^2 \end{aligned} \tag{18}$$

where $\|\,\|_2$ means the Euclidean norm of a vector. Therefore, Eq. (17) is reduced to

$$\boldsymbol{x}_{(m+1)} = \boldsymbol{x}_{(m)} - \gamma\{f(\boldsymbol{x}_{(m)}) - f^*\}[\partial f(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}]^{\mathrm{T}}/\|\partial f(\boldsymbol{x}_{(m)})/\partial \boldsymbol{x}\|_2^2 \tag{19}$$

For example, in an unconstrained non-linear programming problem with two variables, Eq. (19) means that the values of $\boldsymbol{x}_{(m+1)}$ are determined as the point where the line on the tangent of the objective function drawn from the point toward the direction with the steepest descent gradient crosses the plane $f(\boldsymbol{x}) = f^*$, as shown in Fig. 5.

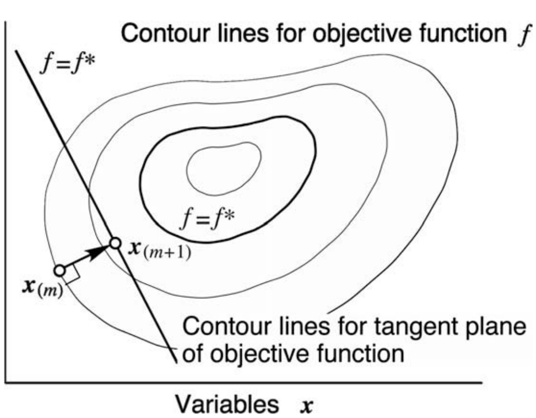

Fig. 5. Renewal of values of variables.

The renewal of the values of the variables $\boldsymbol{x}$ based on the extended Newton–Raphson method has the following features: In the region with a feasible solution, the renewal can have the convergence to it; In the region with no feasible solution, the renewal can create a chaotic behavior and has the capability of global search; In the region with no feasible solution, the renewal can also create a cyclically vibrating behavior and has the possibility of trap into a local optimal solution. To prevent the trap, the decelerating parameter γ is changed randomly in the range $0 < \gamma \leqslant 1$ at each renewal.

5. Numerical studies

The aforementioned neural network model is applied to the prediction of the cooling demand in a building used for a bench mark test of a variety of prediction methods.

5.1. Calculation conditions

The cooling demand is for a commercial building with nine floors, a parking space on the first floor, and a total floor area of 5400 m^2. The cooling demand measured on the 23 weekdays in July, 1996 is used to identify the values of model parameters, or weights and thresholds. The sampling time interval is set at $\Delta t = 1$ h. Therefore, the number of sampling times for the measured cooling demand is $M = 23 \times 24 = 552$. The prediction of the cooling demand is conducted using the identified values of model parameters on the 22 weekdays in August, 1996.

The following numerical studies are conducted:

(a) The effect of the global optimization is investigated by comparing the results obtained by a conventional local optimization method and the modal trimming one. Here, the quasi-Newton method is adopted as a conventional local optimization one, and the Broyden-Fletcher-Goldfarb- Shanno (BFGS) formula is applied to express the inverse of the Hessian matrix approximately. The global optimization by the modal trimming method is also applied in the other numerical studies (b) to (d).
(b) The effect of the periodic differential operation is investigated by comparing the results obtained in the cases with the 1st order periodic differential operation $D = 1$and without any periodic differential operation $D = 0$. Here, the number of sampling times for the periodic differential operation is set at $s = 24$. The 1st order periodic differential operation is also applied in the other numerical studies (a), (c), and (d).
(c) The effect of the number of neurons of the input layer I, which depends on the number of measured or predicted values p and the number of errors between measured and predicted values q, as well as the number of neurons of the hidden layer J is investigated by changing these numbers as $p = 1 \sim 10$, $q = 1 \sim 10$, and $J = 3 \sim 10$. Here, $p = 2$, $q = 1$, and $J = 3$ are selected as reference conditions. These are also used in the other numerical studies (a), (b), and (d).
(d) The effect of the air temperature and relative humidity as supplementary conditions is investigated by comparing the results obtained in the cases where both and none of them are considered. These supplementary conditions are not considered in the other numerical studies (a) to (c).

In all the aforementioned numerical studies, the order of the ordinary difference is set at $d = 0$. The number of sampling times for prediction is set at $L = 3$. The constant coefficient in Eq. (9) is set at $\alpha = 200.0$ in consideration of the magnitude of the periodic difference of the cooling demand.

5.2. Effect of global optimization

First, the quasi-Newton and modal trimming methods are applied to the identification of the values of model parameters. Fig. 6a and b partly compare the measured and predicted cooling demands obtained by the quasi-Newton and modal trimming methods, respectively. Here, the predicted cooling demands are based on the values of model parameters corresponding to a local

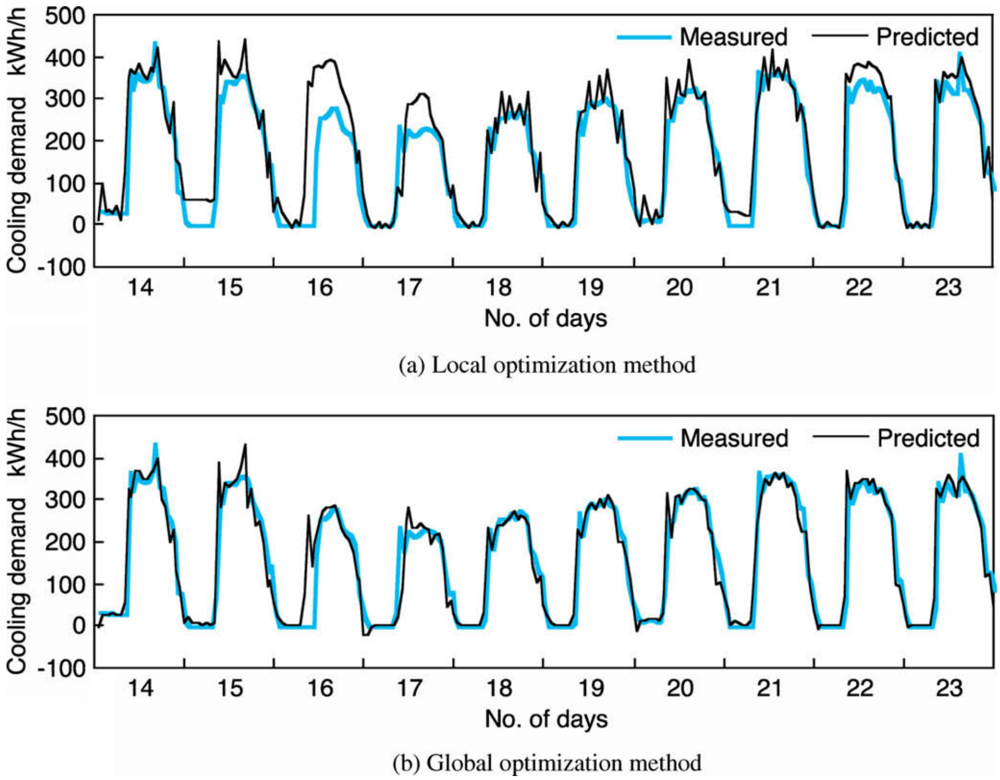

Fig. 6. Comparison of measured and predicted cooling demands for identification obtained by local and global optimization methods.

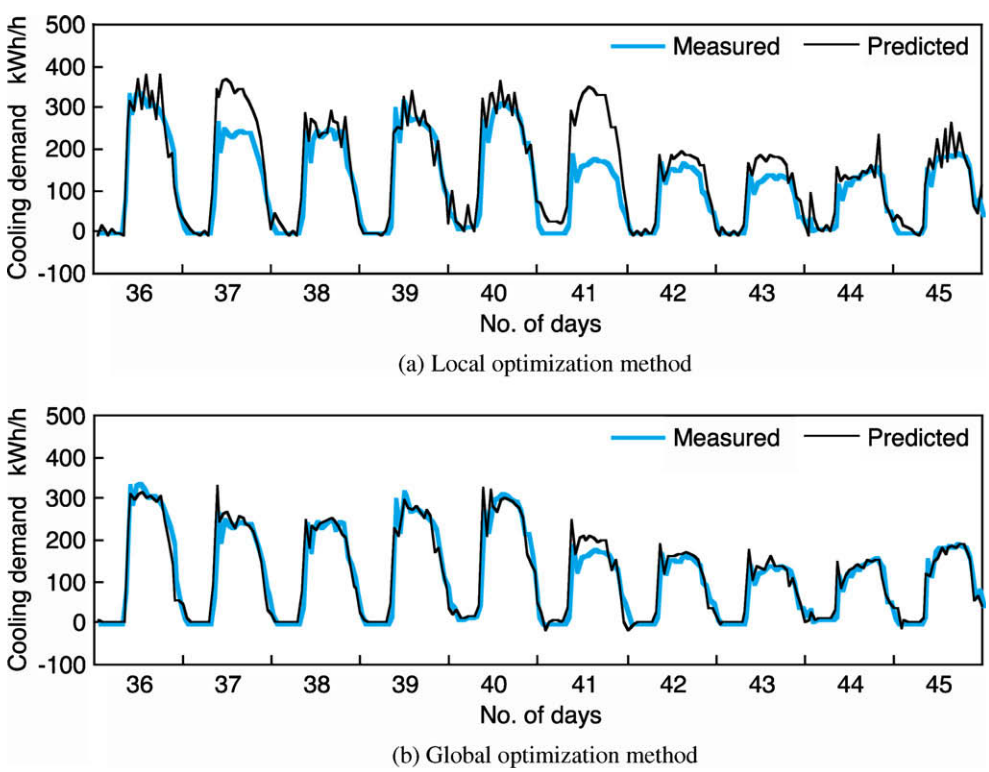

Fig. 7. Comparison of measured and predicted cooling demands for prediction obtained by local and global optimization methods.

optimal solution obtained by the quasi-Newton method and the global quasi-optimal solution obtained by the modal trimming method. The global quasi-optimal solution is obtained through some local optimal solutions. According to Fig. 6a, the predicted cooling demand sometimes fluctuates widely, and the errors between the measured and predicted cooling demands sometimes become too large. Therefore, this result cannot be validated. On the other hand, according to Fig. 6b, the measured and predicted cooling demands are in good agreement. Next, the prediction is conducted using the values of model parameters corresponding to the local optimal solution obtained by the quasi-Newton method and the global quasi-optimal solution obtained by the modal trimming method. Fig. 7a and b partly compare the measured and predicted cooling demands obtained by the quasi-Newton and modal trimming methods, respectively. The results obtained for the prediction are similar to those for the identification. Table 1 shows the absolute and relative values of the root mean square errors for the four cases. Here, each relative value is obtained by dividing the corresponding absolute value by the difference between the maximum and minimum values during the period for the identification or prediction. The values corresponding to the global quasi-optimal solution are much smaller than those corresponding to the local optimal solution for both the identification and prediction.

These results show that the identification of the values of model parameters for the neural network model needs a global optimization method, and that the modal trimming method can find appropriate values as the global quasi-optimal solution.

Table 1
Absolute and relative values of root mean square errors obtained by local and global optimization methods

Case	Identification		Prediction	
	Absolute error (kWh/h)	Relative error (%)	Absolute error (kWh/h)	Relative error (%)
Local optimization	47.39	10.81	43.78	12.47
Global optimization	35.12	8.01	27.83	7.93

5.3. Effect of periodic differential operation

First, the values of model parameters are identified without any periodic differential operation. Fig. 8 partly compares the measured and predicted cooling demands obtained in this case. According to this figure, although the predicted cooling demand does not fluctuate widely, the errors between the measured and predicted cooling demands sometimes become too large, as compared with the result obtained with the 1st order periodic differential operation in Fig. 6b. Therefore, this result cannot be validated. Next, the prediction is conducted using the values of model parameters identified without any periodic differential operation. Fig. 9 partly compares the measured and predicted cooling demands obtained in this case. According to this figure, a similar observation can be found, i.e., the errors between the measured and predicted cooling demands sometimes become too large, as compared with the result obtained with the 1st order periodic differential operation in Fig. 7b. Table 2 shows the absolute and relative values of the root mean square errors for the two cases. The values obtained with the 1st order periodic differential operation shown in Table 1 are much smaller than those obtained without any periodic differential operation shown in Table 2 for both the identification and prediction.

These results show that the preprocessing by the periodic differential operation is necessary to identify the values of model parameters for the neural network model appropriately.

5.4. Effect of numbers of neurons of input and hidden layers

The values of model parameters are identified by changing the number of measured or predicted values p, the number of errors between measured and predicted values q, and the number of neurons of the hidden layer J, and the prediction is conducted using the

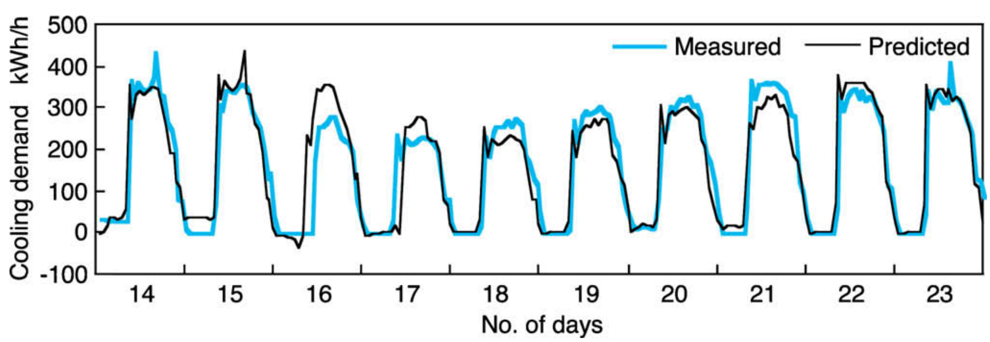

Fig. 8. Comparison of measured and predicted cooling demands for identification obtained without periodic differential operation.

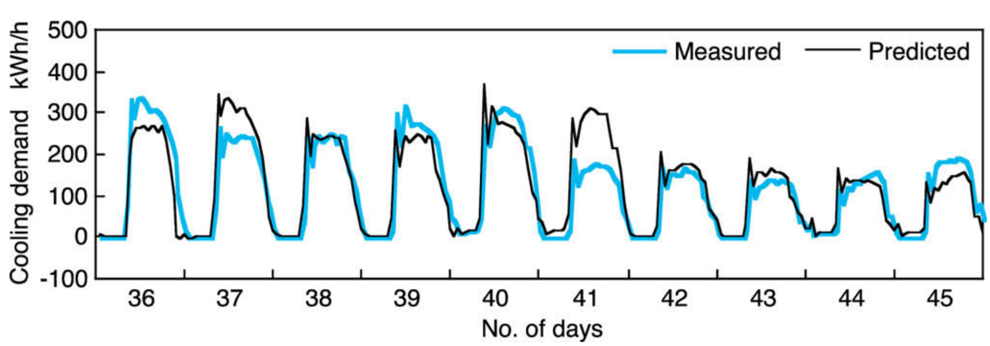

Fig. 9. Comparison of measured and predicted cooling demands for prediction obtained without periodic differential operation.

values of model parameters identified. Table 3 shows the absolute and relative values of the root mean square errors for all the cases. According to this table, the accuracy of the identification of the values of model parameters becomes higher with increases in p and q. On the other hand, the accuracy of the prediction becomes lower conversely with increases in p and q. This is because if the neural network model with a larger number of model parameters becomes excessively adaptable to learned data, then it becomes inadaptable to unlearned data. In this study, J does not significantly affect the accuracy of the identification and prediction.

These results show that the global optimization method can assess the effect of the numbers of neurons of the input and hidden layers on the accuracy of the identification and prediction.

Table 2
Absolute and relative values of root mean square errors obtained without periodic differential operation

Case	Identification		Prediction	
	Absolute error (kWh/h)	Relative error (%)	Absolute error (kWh/h)	Relative error (%)
Without periodic differential operation	40.21	9.17	39.73	11.32

Table 3
Effect of numbers of neurons of input and hidden layers on absolute and relative values of root mean square errors

Number of neurons of input and hidden layers			Identification		Prediction	
p	q	J	Absolute error (kWh/h)	Relative error (%)	Absolute error (kWh/h)	Relative error (%)
1	1	3	35.14	8.02	27.71	7.89
2	1	3	35.12	8.01	27.83	7.92
3	1	3	34.85	7.95	27.96	7.96
5	1	3	34.73	7.92	28.14	8.01
10	1	3	34.07	7.77	28.53	8.13
2	2	3	34.94	7.97	27.95	7.96
2	3	3	34.75	7.93	28.12	8.01
2	5	3	33.69	7.69	29.27	8.34
2	10	3	32.02	7.31	41.32	11.77
2	1	5	35.13	8.02	27.84	7.93
2	1	10	35.03	7.99	27.79	7.92

Table 4
Absolute and relative values of root mean square errors for air temperature, relative humidity, and cooling demand

Predicted quantity	Identification		Prediction	
	Absolute error	Relative error (%)	Absolute error	Relative error (%)
Air temperature	1.34 °C	7.72	1.37 °C	7.12
Relative humidity	6.84%	10.78	6.09%	9.00
Cooling demand	33.44 kWh/h	7.63	28.81 kWh/h	8.20

5.5. Effect of supplemental conditions

First, the values of model parameters are identified independently for the neural network models for predicting air temperature and relative humidity. Next, the prediction of air temperature and relative humidity is conducted independently using the values of model parameters identified. The upper part of Table 4 shows the absolute and relative values of the root mean square errors for air temperature and relative humidity. Second, the values of model parameters are identified for the neural network model for predicting cooling demand by considering predicted values of air temperature and relative humidity as additional inputs. Fig. 10 partly compares the measured and predicted cooling demands obtained in this case. According to this figure, the measured and predicted cooling demands are in good agreement. This result is similar to that shown in Fig. 6b. Finally, the prediction of cooling demand is conducted using the values of model parameters identified. Fig. 11 partly compares the measured and predicted cooling demands obtained in this case. According to this figure, the measured and predicted cooling demands are overall in good agreement. However, there are some differences during nighttime, especially on the 41st day. This is because although the cooling demand becomes zero during nighttime, the air temperature and relative humidity change even during nighttime, and changes in the air temperature and relative humidity affect the prediction of cooling demand. Especially, there is a large drop of the air temperature in the evening on the 40th day, which affects the prediction of cooling demand excessively. The lower part of Table 4 shows the absolute and relative values of the root mean square errors for cooling demand. Because of the aforementioned results, the accuracy of the identification of the values of model parameters becomes higher, while that of the prediction becomes lower.

These results show that the consideration of supplemental conditions can improve the accuracy of the identification of the values of model parameters for the neural network model, but that it may deteriorate the accuracy of the prediction.

6. Conclusions

An approach has been proposed to predict energy demands accurately using neural networks. The trend and periodic change

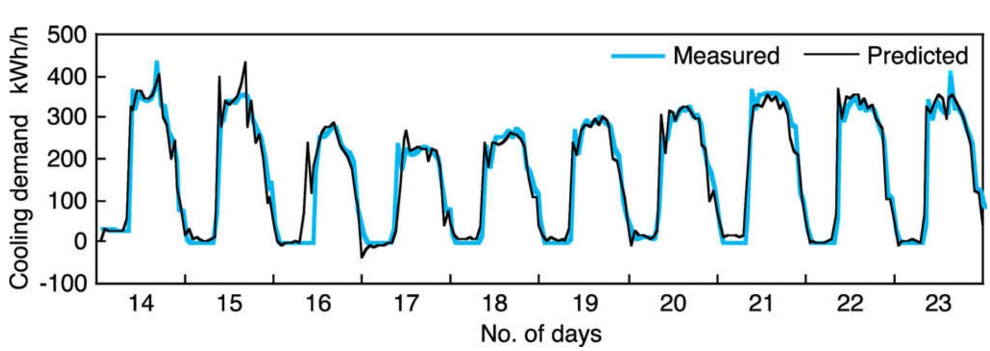

Fig. 10. Comparison of measured and predicted cooling demands for identification obtained with prediction of air temperature and relative humidity.

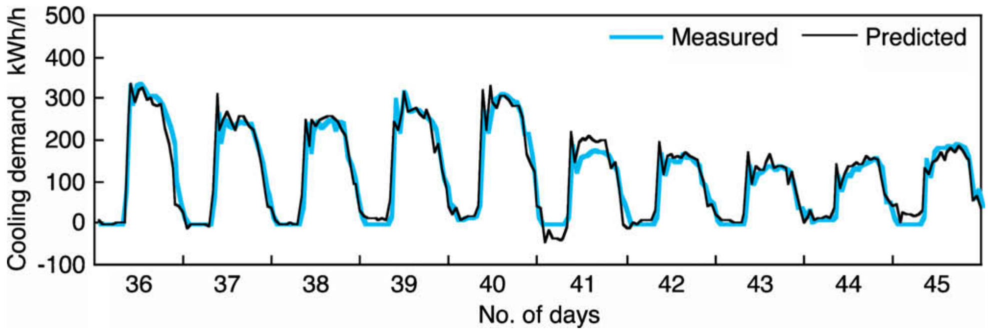

Fig. 11. Comparison of measured and predicted cooling demands for prediction obtained with prediction of air temperature and relative humidity.

have been first removed from time series data on energy demand, and the converted data is used as the main input to a neural network. A global optimization method called "Modal Trimming Method" proposed for non-linear programming problems has been adopted to identify the values of model parameters. In addition, the air temperature and relative humidity have been considered as additional inputs to the neural network, and its effect on the prediction of energy demand has been investigated. This approach has been applied to the prediction of the cooling demand in a building used for a bench mark test of a variety of prediction methods. The results obtained by the proposed approach have been compared with those by the conventional one. The following are the main results obtained:

(1) The identification of the values of model parameters for the neural network model needs a global optimization method, and the modal trimming method can find appropriate values as the global quasi-optimal solution.
(2) The preprocessing by the periodic differential operation is necessary to identify the values of model parameters for the neural network model appropriately.
(3) The global optimization method can assess the effect of the numbers of neurons of the input and hidden layers on the accuracy of the identification and prediction.
(4) The consideration of supplemental conditions can improve the accuracy of the identification of the values of model parameters for the neural network model, but it may deteriorate the accuracy of the prediction.

References

[1] Yokoyama R, Ito K, Kamimura K, Miyasaka F. Development and evaluation of an advisory system for optimal operation of a district heating and cooling plant. In: Proceedings of the international conference on renewable and advanced energy systems for the 21st century, paper no. RAES99-7641, Lahaina; 1999.
[2] Rumelhart DE, McClelland JL, The PDP Research Group. Parallel distributed processing. MIT Press; 1986.
[3] Aydinalp M, Ismet Ugursal V, Fung AS. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. Appl Energy 2002;71(2):87–110.
[4] Senjyu T, Takara H, Uezato K, Funabashi T. One-hour-ahead load forecasting using neural network. IEEE Trans Power Syst 2002;17(1):113–8.
[5] Beccali M, Cellura M, Lo Brano V, Marvuglia A. Forecasting daily urban electric load profiles using artificial neural networks. Energy Convers Manage 2004;45(18–19):2879–900.
[6] Ben-Nakhi AE, Mahmoud MA. Cooling load prediction for buildings using general regression neural networks. Energy Convers Manage 2004;45(13–14):2127–41.
[7] Dodier RH, Henze GP. Statistical analysis of neural networks as applied to building energy prediction. Trans ASME, J Solar Energy Eng 2004;126(1):592–600.
[8] González PA, Zamarreño JM. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. Energy Build 2005;37(6):595–601.
[9] Yang J, Rivard H, Zmeureanu R. On-line building energy prediction using adaptive artificial neural networks. Energy Build 2005;37(12):1250–9.
[10] Yao Y, Lian Z, Hou Z, Liu W. An innovative air-conditioning load forecasting model based on RBF neural network and combined residual error correction. Int J Refrig 2006;29(4):528–38.
[11] Azadeh A, Ghaderi SF, Tarverdian S, Saberi M. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. Appl Math Comput 2007;186(2):1731–41.
[12] Yokoyama R, Ito K. Capability of global search and improvement in modal trimming method for global optimization. JSME Int J, Ser C 2005;48(4):730–7.
[13] Baba N. A new approach for finding the global minimum of error function of neural networks. Neural Networks 1989;2(5):367–73.
[14] Uchida O, Endo T. Minimum search technology via chaotic steepest descent method and its application to back propagation. ISICE Trans Ser A 1995;J78-A(11):1516–8 [in Japanese].
[15] Yokoyama R, Ito K. Global optimization for neural network back propagation by modal trimming method. In: Proceedings of the third China–Japan–Korea joint symposium on optimization of structural and mechanical systems, Kanazawa; 2004. p. 639–44.
[16] Levy AV, Montalvo A. The tunneling algorithm for the global minimization of functions. SIAM J Sci Stat Comput 1985;6(1):15–29.