Combination Model for Short-Term Load Forecasting

Qingming Chen*, Ying Shi and Xiaozhong Xu

School of Information and Electromechanical Engineering, Shanghai Normal University, Shanghai, 0086 / Shanghai, China

Abstract: Gas demand possesses dual property of growing and seasonal fluctuation simultaneously, it makes gas demand variation possess complex nonlinear character. From previous studies know single model for nonlinear problem can't get good results but accurately gas forecast were essential part of an efficient gas system planning and operation. In recent years, lots of scholar put forward combination model to solve complex regression problem. In this paper, a new forecasting model which named regression combined neural network is presented. In this approach we used regression to model the trend and used neural network for calculating predicted values and errors. And to prove the effectiveness of the model, support vector machines(SVM) algorithm was used to compare with the result of combination model. The results show that the combination model is effective and highly accurate in the forecasting of short-term gas load and has advantage than other models.

Keywords: Short term load forecast, regression, detrended data, neural network, SVM.

1. INTRODUCTION

With lack of fossil fuels in our world and increasingly severe environmental pollution problem, we should try to find the green fuel pollution-free or a small amount of contaminated. Natural gas is produced by nature, and they will only produce a small amount of contamination. Natural gas is one of the best choice but has been faced many problems. The most important problem is that we should know how to predict how much gas we will use in next year, next month, next day or even next hour. In the past decades, loading forecast is always a major fields in gas engineering, many generation companies bid for load supply, accurate estimation of short term and medium term demands becomes even more critical. For this purpose, we have been developed a wide variety of approaches for load demand forecasting such as regression analysis, time series analysis, grey theory and artificial neural network and so on. They are based on the ideas of stochastic processes and time series analysis, such as autoregressive moving average (ARMA) [1], seasonal autoregressive integrated moving averages (SARIMA) [2], chaotic time series method [3], multiple regression models [4,5], linear regression models [6,7], exponential smoothing [8],time-varying splines [9], seasonal hybrid nonlinear procedures [10,11], and Markov chain combined model [12]. Some employed soft-computing techniques such as artificial neural network [13,14], artificial intelligence [15], integration of artificial neural network and genetic algorithm [16], fuzzy expert system [17], fuzzy logic [18], self-organizing maps [19], principal component analysis [20], grey system theory [21,22], wavelet transform[23], and support vector regression with chaotic artificial bee colony algorithm [24]. Load demand as a polynomial function of time series has been also used [25]. Taylor *et al.* [26] gives a comparison of four methods for load demand forecasting: SARIMA, exponential smoothing, artificial neural network, and principal component analysis. These methods construct different models, and get few valuable results, and prove that it is highly complicated and is difficult to be modeled and forecast accurately. Each category has its own theoretical limitations and none of them has been identified as being generally superior to others. Such as regression analysis, is belong to statistical model, which needs lots of experimental data, however, in reality, we are lack of experimental data that can be used, it is difficult for us to consider the various factors whose would influence load forecasting and structural model. Time series analysis need experimental data is required is smooth or can be exchange to smooth, so it is difficult to structural model and implement experiment. And artificial neural network as a complex nonlinear system modeling of the important tools, with its good nonlinear characteristics, high fitting precision, flexible and effective learning methods, fully distributed storage structure, is widely used in nonlinear time series forecasting modeling and achieved good effect but it can be result from the influence of the training methods and it often trapped into local minima. In a recent period of time, we find that load forecast is a nonlinear case, and lots of factors can influence and change it. So in reality, now we always adopt a method which is combined linear and nonlinear. In this paper, we propose a novel model-- Combination regression and neural network for short term load forecasting. In this approach, we use regression model for gas demand trend and get residual sequence and the residual sequence is used to train neural network, and the input of artificial neural network include weather, temperature and residual sequence. The output of neural network is the expectation of residual sequence, and using the expectations to revise historical data and get the forecast data of the few next days.

*Address correspondence to this author at the School of Information and Electromechanical Engineering, Shanghai Normal University, Shanghai, 0086 / Shanghai, China;
Email: hellowangchenchen@163.com

1874-4443/13 2013 Bentham Open

We also use this approach to model the seasonal variation of gas demand trend. Using simulation experiments to prove this method is feasibility and validity in the later of this paper. The accuracy of the proposed approach in terms of the maximum absolute error (MAE) and the maximum square error (MSE) is given for load demand analysis of Shanghai. And at the end of this article, the technology of SVM was premised to demonstrate that the model combining regression and neural network for short term load forecasting has advantage than other models.

2. DATA PREPROCESSING

2.1. Data and Data Preprocessing

Through expert experience and known information, we know that the main affect gas load is the weather and temperature, weather is usually divided into eight categories (sunny, cloudy, cloudy to shade, shade to cloudy, shade, shade to rain, rain, snow) and coding them. Temperature includes highest temperature, minimum temperature and average temperature. We use some areas of Shanghai gas historical data (from 2005.11.15 to 2009.11.15) as experimental data, it has 1400 data, 1—1300 is the training data and 1301—1400 is for testing. Gas demand at workday is different from weekend, so we should deal with it. Data has the obvious trend, so before experiments, data should be detrended.

Due to various reasons, historical data will have some abnormal data, so before the experiment judge data is abnormal or not is indispensable. We should take measures to abnormal data into normal or delete some of them. First of all, we should judge data is abnormal or not. If

$$x_t \le \frac{\min\{|m_a|,|m_b|\}}{4} \text{ or } x_t \ge 4*\max\{|m_a|,|m_b|\} \qquad (1)$$

($m_a = median(x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1})$, $m_b = median(x_{t+1}, x_{t+2}\, x_{t+3}\, x_{t+4})$), we consider x_t is abnormal, and then $x_t = \frac{x_{t-2}+x_{t-1}+x_{t+1}+x_{t+2}}{4}$. [27]

The above treatment of data preprocessing is for combination model, and for SVM algorithm, before experiment the data should be normalized to [0,1] as following formula.

$$\hat{x}_i = \frac{x_t - \min(x_i)}{\max(x_i) - \min(x_i)}, \text{ i=1,2,...,n.} \qquad (2)$$

Where $\hat{x}_i$ is the normalized value, $\max(x_i)$ and $\min(x_i)$ are represented the maximum and minimum load in the training sample respectively.

2.2. Gas Demand Curve

Gas demand data of Shanghai from year of 2005 to 2009, we use for experiment of training data and testing data. Gas demand time series should be detrended before experiments: after detrending the load demand time series is in (Fig. **1**). We decompose the historical data into seasonal time series in (Fig. **2**) (include spring, summer, autumn and winter). As gas demand at weekend is different from at workday, we resolve the seasonal time series into workdays and weekend in (Fig. **3**). And irregular component like Spring Festival, National Day and some other special time are in (Fig. **4**).

3. SVM

3.1. SVM Theory

Support vector machine (SVM) is a new and promising technique of data mining for data classification and regression, and have been successfully employed to solve nonlinear regression and time series problems. The SVM method is based on the principle of structural risk minimization (SRM) rather than the principle of empirical risk minimization,

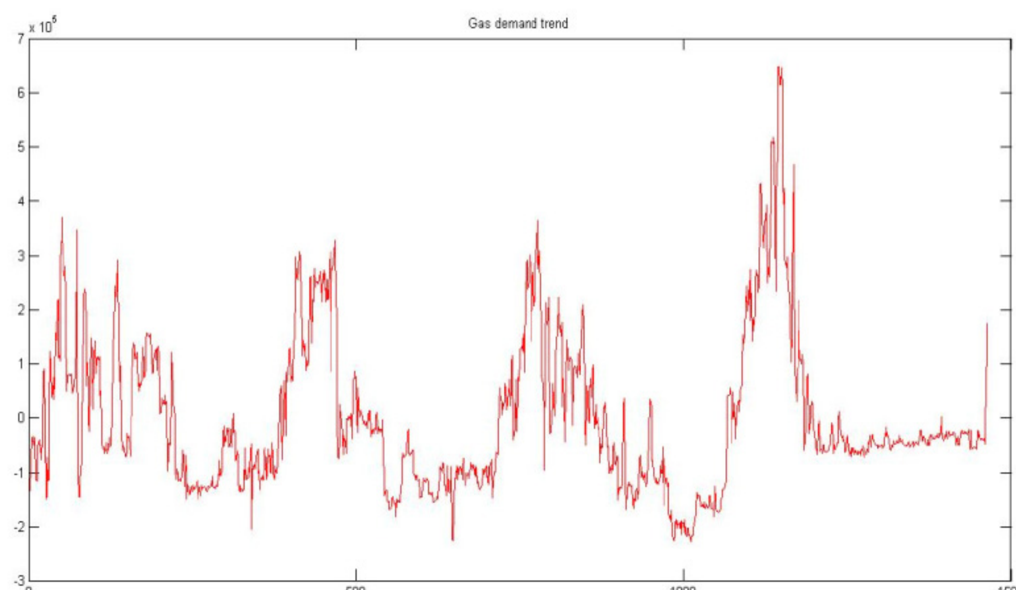

Fig. (1). Gas demand of some areas of Shanghai from 2005.11—2009.11.

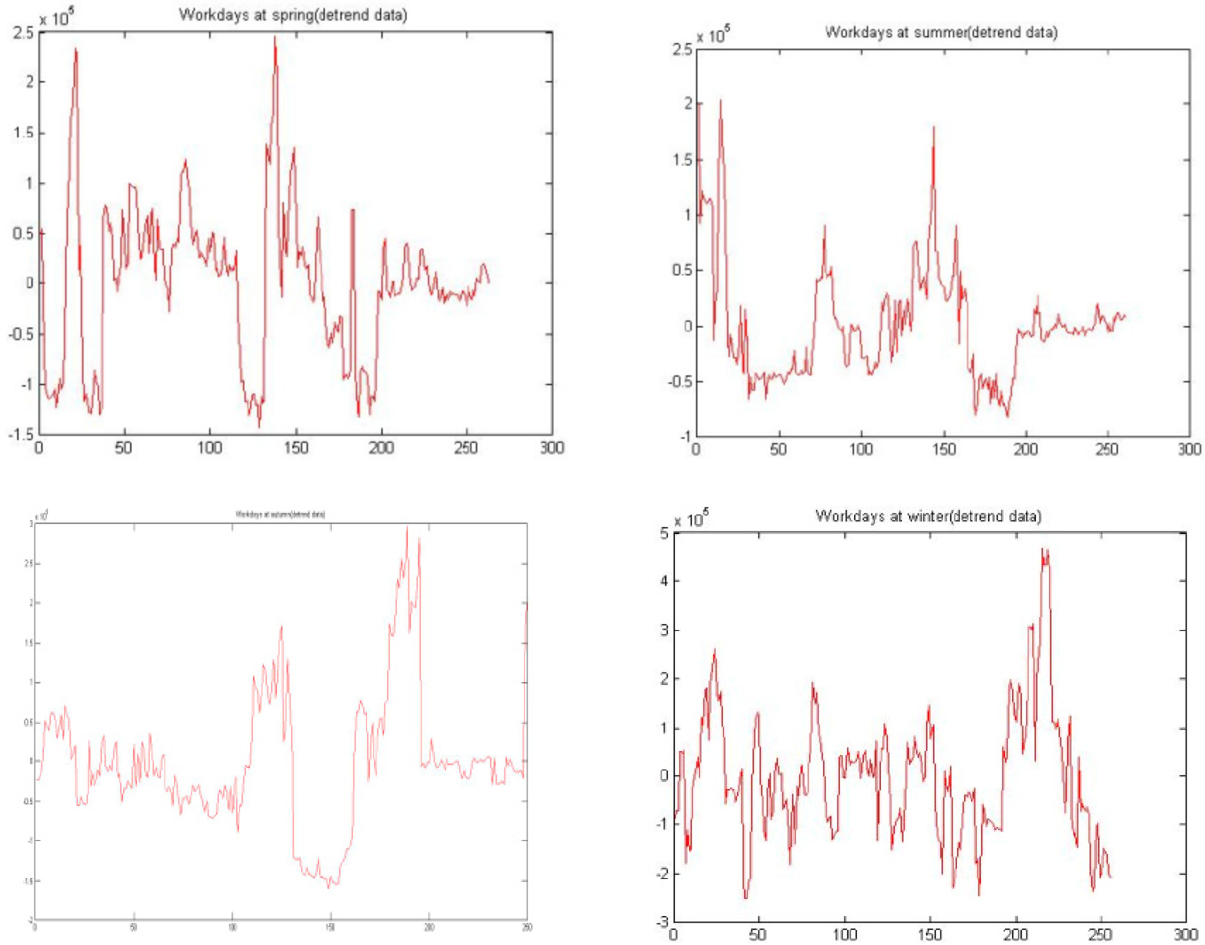

Fig. (2). Gas demand composes into spring, summer, autumn and winter.

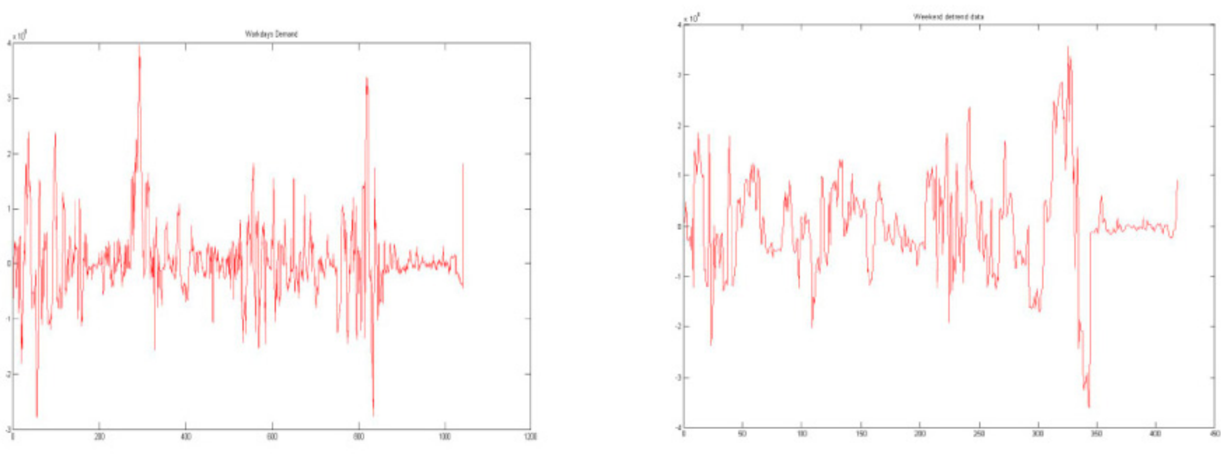

Fig. (3). Gas demand composes into workday and weekend.

which conducted by most of traditional neural network models. With introduction of Vapnik's ε-insensitive loss function [28], SVM has been extended to solve nonlinear regression estimation problems in financial time series forecasting, air quality prediction, production value forecast of machinery industry, engine reliability prediction, etc. At this paper, it is used for forecasting gas demand and that is nonlinear regression problem.

3.2. SVM Regression Theory

In this section, we briefly introduce support vector regression (SVR) which can be used for time series prediction.

Suppose a set of data (x_i, y_i), i=1,2,…,n, $x_i \in R^n$ are given as input, $y_i \in R$ are the corresponding output. SVM regression theory is to find a nonlinear map from input space to output space and map the data to a higher dimensional feature space through the map, then the following estimate function is used to make linear regression [28,29]. The support vector regression solves an optimization problem:

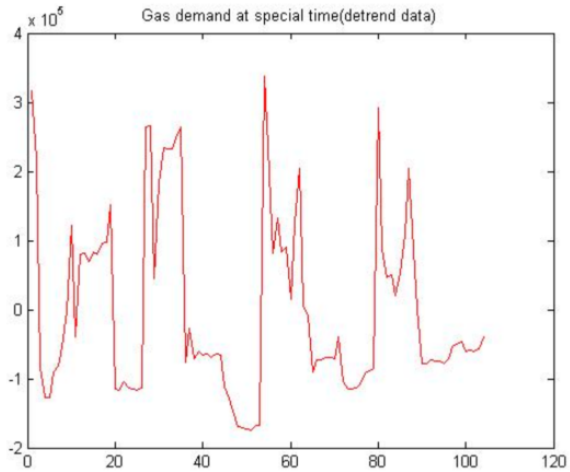

Fig. (4). Gas demand at special time (such as Spring Festival.

$$\min_{w,b,\xi,\xi^*}\{\frac{1}{2}w^T w + C\sum_{i=1}^{l}(\xi_i+\xi_i^*)\} \quad (3)$$

Subject to $y_i-(w^T\phi(x_i+b))\le\varepsilon+\xi_i$,

$(w^T\phi(x_i)+b)-y\le\varepsilon+\xi_i^*$

$\xi_i,\xi_i^*\ge 0$

$i=1,2,...,l$

Where x_i is mapped into a higher dimensional space by the function ϕ . ξ_i is the upper training error (ξ_i^* is the lower) subject to the ε-insensitive tube $|y_i-(w^T\phi(x_i)+b)|\le\varepsilon$. The parameters which control the regression quality are the cost of C, the width of the tube ε and the mapping function ϕ .

The constraints of (11) imply that we would like to put most data x_i in the tube $|y_i-(w^T\phi(x_i)+b)|\le\varepsilon$. If x_i is not in the tube, there is an error ξ_i or ξ_i^* which we would like to minimize in the objective function. SVR avoids underestimating or overestimating the training data by minimizing the training error $C\sum_{i=1}^{l}(\xi_i+\xi_i^*)$ as well as the regularization term $\frac{1}{2}w^T w$. For traditional least-square regression ε is always zero and data are not mapped into higher dimensional spaces. With the Lagrange multipliers introduced, the decision function given in (3) can be expressed as the following explicit form:

$$f(x)=\sum_{i=1}^{l}(a_i-a_i^*)k(x_i,x_j)+b \quad (4)$$

Where a_i,a_i^* are Lagrange multipliers with $a_i\times a_i^*=0$ and $a_i,a_i^*\ge 0$ for any i=1,2,…, l .

Using Mercer's theorem, the regression is obtained by solving a finite dimensional QP problem in the dual space avoiding explicit knowledge of the high dimensional mapping and using only the related kernel function. In (4), we introduced a kernel function $K(x_i-x_j)=\phi(x_i)\times\phi(x_j)$, which is the inner product of two vectors in feature space $\phi(x_i)$ and $\phi(x_j)$. It can be shown that and symmetric kernel function K satisfying Mercer's condition corresponds to product in some feature space. A common kernel function is RBF kernel adopted in the paper which is shown as follows.

$$K(x,y)=\exp(-\|x-y\|^2/(2\sigma^2)) \quad (5)$$

Thus, the Lagrange multipliers can be obtained by maximizing the following form:

$$R(a_i,a_i^*)=-\frac{1}{2}\sum_{i,j}^{n}(a_i-a_i^*)(a_j-a_j^*)K(x_i,x_j)-\varepsilon\sum_{i=1}^{l}(a_i+a_i^*)+\sum_{i=1}^{l}y_i(a_i-a_i^*) \quad (6)$$

Subject to:

$\sum_{i=1}^{n}a_i=\sum_{i=1}^{n}a_i^*, \sum_{i=1}^{n}a_j=\sum_{i=1}^{n}a_j^*$ $0\le a_i\le C, i=1,2,...,l$,

$0\le a_i^*\le C, i=1,2,...,l$,

$0\le a_j\le C, j=1,2,...,l$,

$0\le a_j^*\le C, i=1,2,...,l$ (7)

Through adjusting the two parameters C and ε , the generalized performance can be controlled in high-dimension space. According to Karush-Kuhn-Tucker (KKT) conditions, only some of coefficients $(a_i-a_i^*)$ and $(a_j-a_j^*)$ in differ from zero, and the corresponding training data are referred to as support vector, which can be regarded as the number of neurons in hidden layer of the network structure. Hence SVR is a general and flexible treatment on regression problems and through results to demonstrate which one is better to deal with short-term gas forecasting.

3.3. Training Algorithm of SVM

Sequential minimal optimization (SMO) is a very efficient method for training support vector machines (SVM). SMO can overcome the disadvantage of requiring extremely large memory for storing the kernel matrix and find the solution of the problem efficiently [30].

3.4. Parameters Selection

The performance of SVM regression is affected by parameter selection to some extent. The method of self-adaptively adjusted parameter is adopted in this paper, so selecting optimal parameters by miscellaneous steps can be avoided.

Insensitive parameter ε , penalty factor C and kernel function's width parameter σ need to be determined in SVM of Gaussian kernel function. Insensitive parameter ε is proportional to the noise level of samples. Based on the empirical tuning, its expression is given in (8) [31]:

$$\varepsilon = 3\gamma\sqrt{\frac{\ln(n)}{n}} \tag{8}$$

Where γ denotes the standard deviation of noise, n is the capacity of sample.

Using training algorithm of SVM, penalty factor C and kernel function's width parameter σ can be adjusted by controlling the precision of relative error. Relative error is given as follows.

$$E = \frac{1}{2}\sum\frac{|y_i - f(x_i)|}{y_i} \tag{9}$$

Where y_i is actual loads and $f(x_i)$ denotes forecasting loads.

3.5. Basic Procedures of Load Forecasting

1. Preprocess historical load data, and select pertinent sample from historical data as samples.
2. Select appropriate input variables, mainly including the past n-day load demand data, previous one-day load data, day type, humidity, average temperature, maximum temperature and minimum temperature.
3. Use adaptive SVM algorithm to train sample set to estimating function of regression.
4. Use estimating function of regression to forecast the load of some period in the future.

From simulation of SVM we can get some results in (Table **1**).

$$error = \frac{f(x_i) - y_i}{y_i}\% \tag{10}$$

Table 1. Results from SVM

Data number	Actual value	Predicted value	Error
1301	106800	123154.62	15.3%
1302	107800	108164.06	0.33%
1303	109300	108505.23	-0.7%
1304	110000	103007.06	-6.35%
1305	108100	106008.38	-1.93%
1316	108800	112871.50	3.74%
1317	110900	103108.76	-7.02%

4. REGRESSION AND ANN MODEL

4.1. Demand Analysis for Some Areas of Shanghai

In this paper, we use some areas of Shanghai gas historical data (2005--2009) as experiment data. As declared above, the load demand at weekends is lower than at weekdays, so it needs to have different gas demand curve. And it is obvious that different season has different gas demand, we can get several different gas demand curve. Thus for this analysis we consider four seasons in a year which should delete the special time such as Spring Festival and National Day and others as at Spring Festival always has special value which is different from the usual value, and it can lead to the accurately of the forecast results decreased, so at special time, we should special treatment.

4.2. Regression Model

Due to load demand growth trend is not a simple linear regression, and it presents a nonlinear characteristics. First of all we need to adopt a quadratic regression curve to model its growth trend.

If the historical data series is $Y t_i$ (t=1,2,3,…T , i=1,2,3,4),building a nonlinear regression equation:

$$Y t_i = b_0 + b_1 t_1 + b_2 t_2 + b_3 t_3 + b_4 t_4 + \varepsilon \tag{11}$$

b_0, b_1, b_2, b_3, b_4 and ε is constant.

Using the normally least squares to estimate the model parameters, and get the estimate of parameters, and then the equation of load demand trend is ,

$$\hat{Y} t_i = \hat{b}_0 + \hat{b}_1 t_1 + \hat{b}_2 t_2 + \hat{b}_3 t_3 + \hat{b}_4 t_4 + \varepsilon \tag{12}$$

, and residual sequence is the forecasted data subtract historical data

$$\varepsilon_t = \hat{Y} t_i - Y t_i \quad t=1,2,\ldots,T, i=1,2,3,4, \tag{13}$$

and ε_t is used for training ANN.

$$Y_t = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & x_{11} \cdots & x_{1p} \\ 1 & x_{21} \cdots & x_{2p} \\ \cdots & \cdots\cdots & \cdots \\ 1 & x_{n1} \cdots & x_{np} \end{pmatrix}; \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

So equation (11) transformed into

$$Y = XB + \varepsilon \tag{14}$$

$$X'X = \begin{pmatrix} 1 & 1\cdots & \cdots 1 \\ x_{11} & x_{21}\cdots & \cdots x_{n1} \\ \vdots & \vdots & \vdots \\ x_{1p} & x_{2p} & x_{np} \end{pmatrix}\begin{pmatrix} 1 & x_{11} \cdots & x_{1p} \\ 1 & x_{21} \cdots & x_{2p} \\ \cdots & \cdots\cdots & \cdots \\ 1 & x_{n1} \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n x_{i1} \cdots & \sum_{i=1}^n x_{ip} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 \cdots & \sum_{i=1}^n x_{i1} x_{ip} \\ \cdots & \cdots & \\ \sum_{i=1}^n x_{ip} & \sum_{i=1}^n x_{ip} x_{i1} \cdots & \sum_{i=1}^n x_{ip}^2 \end{pmatrix} \tag{15}$$

$$X'Y = \begin{pmatrix} 1 & 1\cdots & 1 \\ x_{11} & x_{21}\cdots & x_{n1} \\ \vdots & \vdots & \vdots \\ x_{1p} & x_{2p} & x_{np} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1} y_i \\ \sum_{i=1}^n x_{ip} y_i \end{pmatrix} \tag{16}$$

$$X'X\hat{B} = X'Y \tag{17}$$

$$\hat{B}=\begin{pmatrix}\hat{b}_0\\ \hat{b}_1\\ \vdots\\ \hat{b}_p\end{pmatrix}=(X^{'}X)^{-1}X^{'}Y \tag{18}$$

$$\hat{y}=\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2+\cdots+\hat{b}_px_p \tag{19}$$

4.3. Results from Regression

Through calculation gets results and some results show in (Table **2**).

Table 2. Results from Regression

Data number	Actual value	Predicted value	Error
1301	106800	123164.32	15.3%
1302	107800	108264.08	-0.43%
1303	109300	108555.63	0.68%
1304	110000	103087.44	-6.23%
1305	108100	105908.73	-2%
1316	108800	113001.50	3.87%
1317	110900	103308.36	-6.8%

4.4. ANN Model

Artificial neural networks are designed to simulate the biological nervous system, which are based on simulated nerve neurons which are joined together in a variety of ways to form networks. Before we use network, we should first determine its structure and weights by training. Training a neural network is to find the optimal function by adjusting the values of the weights between elements. Most widely used artificial neural network is because of these network have the capacity to learn, memorize and create relationships amongst data, and multilayer perceptron models is mostly used, in MLP models there is always one input neuron, with a number of neurons equal to the number of variables of the problem, and one output neuron. The perceptron response is made available, with a number of neurons equal to the desired number of quantities computed from the inputs. In neural network, BP neural network with three layers is most commonly used, a three layers BP neural network can approach arbitrary precision any given continuous functions. Now we often use Levenberg—Marquardt algorithm to train BP neural network. Levenberg—Marquardt algorithm is one of the best algorithm which is the fastest convergence and robustness. The general procedure of this algorithm mainly includes data preprocessing, training network and simulation.

4.5. Determining Network Structure

Data from 15 November, 2005 to 15 November, 2009 is offered by a gas company in some area of Shanghai, which is used for training and testing. So the historical data is the most important role in the prediction. And through lots of study found that there are two basic factors which affect gas demand in the short term (i.e. the current day and the following day). There are weather and customer behavioral patterns.

4.5.1. Weather

A wide variety of temperature statistics are used to predict gas demand. We assume that gas load is not only affect by daily temperature but also for last 2 days, last week and last month previous to the day. It is obvious that temperature include maximum temperature, minimum temperature and average temperature. These temperatures are usually plotted against the gas load. In order to get accurate results, so the exact values of temperature are indispensable.

4.5.2. Customer Behavioral Patterns

Current behavioral patterns strongly relate to gas demand because most people work during the period 9 a.m. to 5 p.m. Mondays to Fridays, whilst they are at leisure during the weekend. Gas demand when people are working is higher than people they are rest at home. During holiday, people hardly use gas. So for convenience, these variations in customer behavior can be sub-divided into the effects produced by day of week and holidays.

4.5.2.1. Day of Week

As be expected, daily demand varies between a high day time level and a low night time level. In addition, due to the behavior of the industrial sector of the market, these levels of sales are higher on weekdays than on weekend.

During the weekdays, they are marked “breakfast” (7:00 to 8:30 a.m.) and “tea-time” (5:00 p.m. to 6:30 p.m.), and in that time, gas demand are at peaks. On weekends the breakfast peak is later and less pronounced and the “tea-time” peak is often proportionally higher than during the workdays.

4.5.2.2. Holidays

At holiday times, the gas demand is lower than normal for that time of year, due to the partial or complete shutdown of industry. The effect of gas demand on holiday period at Spring Festival and National Day is particularly difficult to estimate. When Spring Festival falls on a weekday it is extremely significant. If Spring Festival is closed to a week (either side) it will extend the holiday period to include this weekend. This may result in industry completely shutting down. National Day effects gas demand as well. So customer behavioral patterns have a major impact on gas forecasting

From discussed above, we assume that the number of input layer neuron is 5, and only need one output. To determine the number of hidden neurons is related to the precision of the BP network and learning efficiency, at present it's sure there is no general guiding principle, and just only can according to some rules of thumb or through test to determine it. Such as empirical formula: $N_h=\sqrt{N_o+N_i}+N_i$ Hecht—Nielsen put forward that $N_h=2N_i+1$, N_i is the number of input nodes, N_h is the number of hidden nodes, N_o is the number of output nodes, and N_j is between 1—10. By changing the constant N_j, different numbers of hidden neurons are obtained. Table **3** shows the different hidden neurons of the test results.

Table 3. Different Hidden Neurons of the Test Results

Hidden node number	MAE (%)	MSE (%)
3	3.3823	3.990
4	3.1900	4.7603
5	3.4233	3.7690
6	3.4700	3.8254
7	3.3300	3.7047
8	3.4043	3.9159
9	3.4655	3.9503
10	3.4572	4.0132
11	3.2744	3.9766
12	3.5061	3.4701

According to Table **1**, because of 5-11-1 type has the best performance, this type is chosen for the BP neural network (BPNN) forecasting method.

In this paper we use empirical formula to determine the number of hidden nodes and first assume it is 11. So the structure of network in this paper is BP (5, 11, 1). As discussed above, input includes actual data sequence, maximum temperature, minimum temperature, average temperature and week parameters (workday and weekend). And the simulation figure is shown in (Fig. **5**).

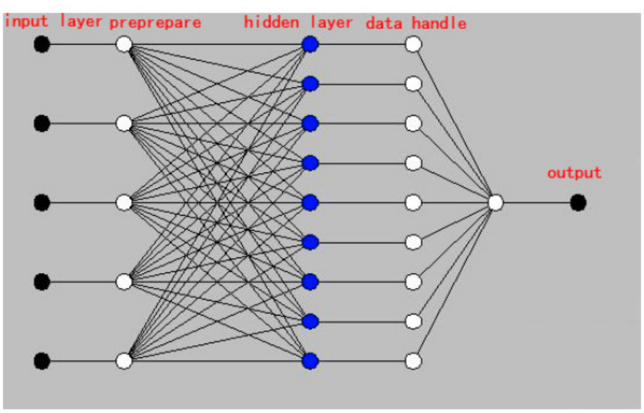

Fig. (5). Structure of neural network.

4.6. Training Method

We use matlab to simulate BP neural network. The hidden layer uses tansig as transfer function and purelin as the output layer function. Training method of the network is trainlm-(Levenberg-Marquardt). Due to the interlayer activation function chooses is type of 'S', so before simulation training data should be normalized processing and made the data normalized to [0,1]:

$$y=\frac{x-\min(x)}{\max(x)-\min(x)} \tag{20}$$

4.7. Result from BP Neural Network

Through composing data into four seasons, weekends and workdays and special days, we can get results from BP neural network. And some experiment results are in (Table **4**).

Table 4. Result from BP Neural Network

Data number	Actual value	Predicted data	Error
1301	106800	123444.32	5.16%
1302	107800	108264.08	1.68%
1303	109300	108555.63	-8.5%
1304	110000	104087.36	-3.1%
1305	108100	115908.73	-5.9%
1306	108800	113001.50	1.35%
1307	110900	103349.44	5.38%

From the above form, we can see that the experiment errors is relative bigger, prediction accuracy is not high. So we should combination regression and BP neural network to forecast gas demand.

5. COMBINATION EXPERIMENT AND RESULTS

Though regression and BP neural network can be one of the algorithms to forecast gas demand, from Table **2** and Table **3** we can see that both of them have high error rate, but accurately gas forecast were essential part of an efficient gas system planning and operation. If the results have not high accuracy, we should try to choose another way and make sure to get good results. In recent years, most of scholars put forward that combine linear model and nonlinear model to forecast time series problems. In this paper, we combine regression and BP neural network to forecast gas demand and experimental steps are in below.

1. Using the method of regression analysis data, getting demand trend chart and calculating residual sequence ε_i.
2. Using BP neural network to train the residual sequence, and through training network to get residual expectations sequence $\hat{\varepsilon}_i$.
3. Using expectations sequence to revise residual sequence, and then get predicted value of the next days.
4. Calculating errors from this combination model and get the chart of gas demand.

From the experiment steps, the data preprocessing before experiment is the first thing we should do. As the same of BP neural network (BPNN), the structure of combination model should be determined as well. The basic procedure of the combination is as following figure (Fig. **6**).

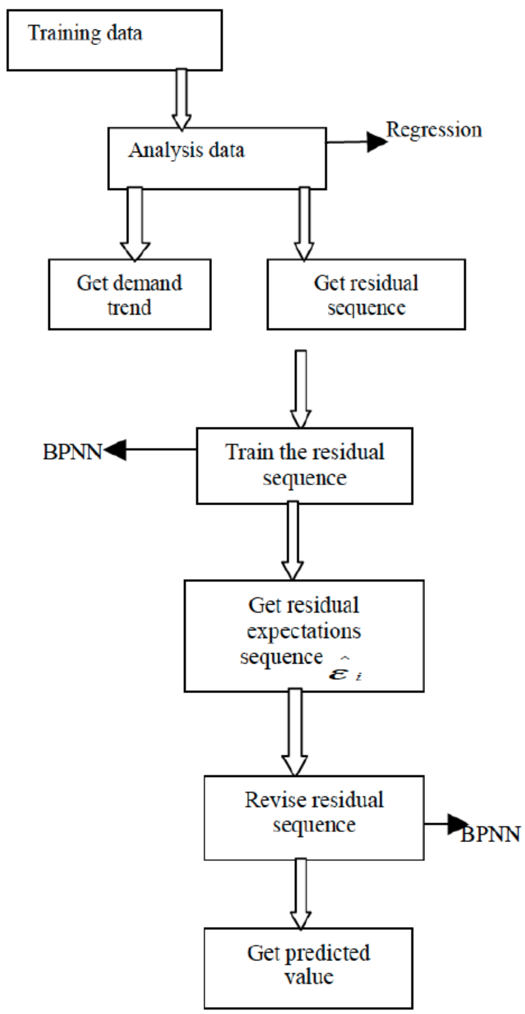

Fig. (6). Basic procedure of combination model.

Through combining regression and BP neural network we can get some results and some of them are at Table **5**.

From Table **2**, Table **4** and Table **5**, error from combination experiments is smaller than regression model and BP neural network. Next section we use MAE and MSE to state that combine regression and BP neural network is effective and feasible for load forecasting and can get more accurate results.

6. ANALYSIS AND CONCLUSIONS

As mentioned above, in this paper the experiment combination nonlinear regression and neural network for short--term load forecasting, and then use SVM and BP neural network to illustrate that apply this approach can get more accurate results. SVM experimental diagram is (Fig. **7**). In this section, employ mean absolute errors (MAE) and mean square errors (MSE) to specify the experiment results.

Table 5. Results from Combination Model

Data number	Actual value	Combination experiment	Error
1301	106800	108299.7	1.40%
1302	107800	109235.1	1.33%
1303	109300	104290.7	-4.5%
1304	110000	106251.2	-3.4%
1305	108100	110334.7	-2.9%
1306	108800	109880.1	0.9%
1307	110900	112245.7	1.8%

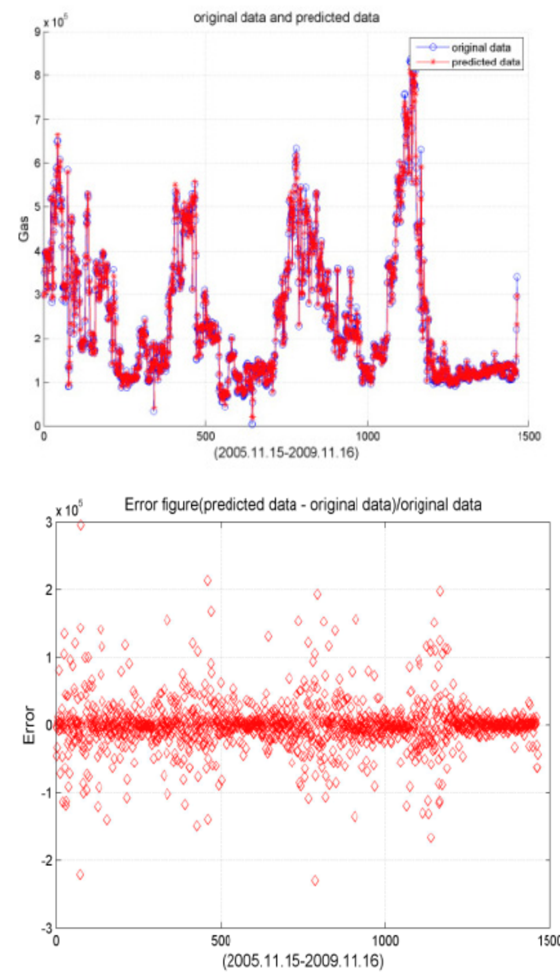

Fig. (7). Gas demand curve of SVM.

The mean absolute errors (MAE) of the predictions were in general lower for the combination regression and ANN than for the SVM and BP neural network.

For each experiment the mean square errors (MSE), the values of MSE from SVM is lower than from combination model and BP neural network, but SVM need more time to train for the same historical data. The results of the same experimental data are in (Table **6**).

Table 6. MAE and MSE from Experiment

	MAE	MSE(10^4)
SVM	0.95	8.299
regression	1.79	6.757
BP neural network	3.2744	3.9766
Combination model	1.36	3.7058

From the above experiment, the MAE and MAPE from the approach of combination nonlinear regression and neural network are lower than SVM and BP neural network. And we can see the approach of combination neural network and neural network is effective and can get more accurately result.

From this paper we can see that forecast load demand accurately is a difficult task and has several influencing factors which influence the demand result and these factors can't be forecasted accurately by only one method. So in order to get accurate result, it is usually adopt the method of combination linear and nonlinear to forecast the load demand.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENT

This work was supported by Science And Technology Commission Of Shanghai Municipality, Shanghai, China (Grant No.11510502400).

REFERENCES

[1] L. Magnano and J.W Boland, "Generation of synthetic sequences of electricity demand: application in South Australia", *Energy*, vol. 32, p. 2230e43, 2007.

[2] L.J Soares and L.R Souza, "Forecasting electricity demand using generalized long memory" *Int. J. Forecast*, vol. 22, no. 17, p.28, 2006.

[3] J.Z. Wang, D.Z. Chi, J. Wu and H.Y. Lu, "Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting", *Expert Syst. Appl.*, vol. 38, no.7, p. 8419e29, 2011.

[4] R. Ramanathan, R. Engle, C.W.J Granger, F. Vahid-Araghi and C. Brace, "Short-run forecasts of electricity loads and peaks", *Int. J. Forecast*, vol. 13, p. 161e74, 1997.

[5] M.J Thatcher, "Modelling changes to electricity demand load duration curves as a consequence of predicted climate change for Australia", *Energy*, vol. 32, p. 1647e59, 2007

[6] V. Bianco, O. Manca and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models", *Energy*, vol. 34, no.9, p. 1413e21, 2009.

[7] A. Goia, C. May and G.F Goia, "Functional clustering and linear regression for peak load forecasting", *Int. J. Forecast*, vol. 26, no. 4, p. 700e11, 2010.

[8] J. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing", *J. Oper. Res. Soc.*, vol. 54, p. 799e805, 2003.

[9] A. Harvey and S.J Koopman, "Forecasting hourly electricity demand using timevarying splines", *J. Am. Stat. Assoc.*, vol. 88, p. 1228e36, 1993.

[10] H.T Pao, "Forecasting energy consumption in Taiwan using hybrid nonlinear models", *Energy*, vol. 34, no.10, p. 1438e46, 2009.

[11] S. Zhu, J. Wang, W. Zhao and J. Wang, "A seasonal hybrid procedure for electricity demand forecasting in China", *Appl. Energy*, vol. 88, p. 3807e15, 2011.

[12] H. Wang, S. Zhu, J. Zhao and G. Li, "An improved combined model for the electricity demand forecasting", *Int. Conf. Comput. Inf. Sci.,* vol. 108e11. p. IEEEXplore, 2010.

[13] A. Azadeh, S.F Ghaderi and S. Sohrabkhani, "A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran", *Energy Policy*, vol. 36, no.7, p. 2637e44, 2008.

[14] P.C Chang, C.Y Fan and J.J Lin, "Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach", *Int. J. Electr. Power Energ. Syst.*, vol. 33, no. 1, p. 17e27, 2011.

[15] K. Metaxiotis, A. Kagiannas, D. Askounis and J. Psarras, "Artificial intelligence in short term electrical load forecasting: a state-of-the-art survey for the researcher", *Energy Convers. Manage.*, vol. 44, p. 1525e34, 2003.

[16] A. Azadeh, F. Ghaderi and S. Tarvardiyan, "Integration of artificial neural network and genetic algorithm to predict electricity energy consumption", *Appl. Math. Comput.*, vol. 186, no. 2, p. 1731e41, 2007.

[17] P.K Dash, A.C Liew, S. Rahman and G. Ramakrishna,"Building a fuzzy expert system for electric load forecasting using a hybrid neural network", *Expert Syst. Appl.,* vol. 9, no.3, p. 407e21, 1995.

[18] S. Kucukali and K. Baris, "Turkey's short-term gross annual electricity demand forecast by fuzzy logic approach", *Energy Policy*, vol. 38, no. 5, p. 2438e45, 2010.

[19] Lendasse A, Lee J, Wertz V and Verleysen M, "Forecasting electricity consumption using nonlinear projection and self-organizing maps", *Neuro. Comput.,* vol. 48, no. 1e4, p. 299e311, 2002.

[20] M. Manera and A. Marzullo, "Modelling the load curve of aggregate electricity consumption using principal components", *Environ. Model Softw.*, vol. 20, p. 1389e400, 2005.

[21] J.L. Deng, "Introduction to grey systems", *J. Grey. Syst.*, vol. 1, p. 1e24, 1989.

[22] P. Zhou, B.W. Ang and K.L. Poh, "A trigonometric grey prediction approach to forecasting electricity demand", *Energy*, vol. 31, no.14, p. 2839e47, 2006.

[23] H.T Nguyen and I.T Nabney, "Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models", *Energy*, vol. 35, no. 9, p. 3674e85, 2010.

[24] W-C Hong, "Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm", *Energy,* vol. 36, no. 9, p. 5568e78, 2011.

[25] I.S Qamber, Maximum annual load estimation and network strengthening for the kingdom of Bahrain, *Arab. Gulf J. Sci. Res.*, vol. 28, no.4, p. 14e23, 2010.

[26] J.W Taylor, L.M Menezes and P.E McSharry, "A comparison of univariate methods for forecasting electricity demand up to a day ahead", *Int. J. Forecast,* vol. 22, no.1, p. 1e16, 2006.

[27] M. Jin, X. Zhou, Z.M. Zhang and M.M. Tentzeris, "Short-term power load forecasting using grey correlation contest modeling", *Expert Syst. Appl.*, vol. 39, pp. 773-779, 2012.

[28] Z. Liang, L. Wang, and D. Fu, "Short-term power load forecasting based on Lyapunov", *Proc. CSEE*, vol. 18, pp. 368-472, 1998 .

[29] V. N. Vapnik and Xucgong Zhang, "*Nature of Statistics Theory*", Tsinghua University Press, Bejing, 2000.

[30] S.K. Shevade, S.S. Keerthi and C. Bhattacharyy, "Improvements to SMO algorithm for SVM regression", *IEEE Trans*, vol 11, no. 5, pp.1188-1193, 2000.

[31] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression", *Neural. Netw.*, vol. 17, no.1, pp. 113-126, 2004.

Received: August 13, 2013 Revised: August 28, 2013 Accepted: August 28, 2013