Applied Mathematics & Information Sciences
An International Journal

Short-term Load Forecasting of Smart Grid Systems by Combination of General Regression Neural Network and Least Squares-Support Vector Machine Algorithm Optimized by Harmony Search Algorithm Method

Ming Zeng[1], *Song Xue*[1], *Zhijie Wang*[1,*], *Xiaoli Zhu*[1], *Ge Zhang*[1]

[1]Research Advisory Center of Energy and Electricity Economic , North China Electric Power University, Beijing 102206 P. R. China

Received: 7 Jun. 2012, Revised: 21 Sep. 2012, Accepted: 29 Sep. 2012
Published online: 1 Feb. 2013

Abstract: This paper presents an optimization algorithm to solve the short-term load forecasting problem more quickly and accurately in progress of smart grid development. The new approach employs generalized regression neural network (GRNN) to select influence factors of short-term load, and then a least squares-support vector machine (LS-SVM) based on harmony search algorithm (HS) optimization algorithm was proposed that improving the computing accuracy and speed through a novel category of bionic algorithm, and determining the hyper-parameters of LS-SVM through HS optimization algorithm fleetly and reasonably. Simulations have been made comparing the proposed algorithm with several other algorithms commonly used to solve short-term load forecasting problems. The actual implementation result proves that the proposed algorithm can achieve higher prediction accuracy and better computational speed which is more practical for short term load forecasting.

Keywords: Short-term load forecasting, Generalized regression neural network (GRNN), Hyper-parameters selection, Harmony search algorithm (HS), Least squares-support vector machine (LS-SVM).

1 Introduction

During the Twelfth Five-Year planning period (Five-Year Plan is the most important government document of country-regionplaceChina. It is a series of economic development initiatives, mapping strategies for economic development, setting growth targets and launching reforms in relative time frame.), large-scale construction of smart grid and ensuring the security and stability operation of smart grid require improving load forecasting methods and achieve short-term load forecasting timely and accurately [1,3]. The operation of smart grid needs more timely and accurate load forecasting to provide decision support for dispatching and load managing. To satisfy smart grid operation requirements about load forecasting speed and accuracy, timely and accurate load forecasting methods are in great need of. In general, along with the full-scale smart grid construction, the power supply mode and consumption mode of the whole system can be optimized through accurate short-term load forecasting. The security, stability and cleanness of the system are also enhanced further. Therefore, achieving short-term load forecasting is objective requirement of smart grid construction and has become the focus of market subjects and is also one of important issues need to be deepened and broken through.

Load forecasting has become a crucial issue for the markets subjects and researchers of electric power systems. For a long time, most of short-term load forecasting approaches are based on time series analysis method and statistical model, such as linear regression methods [4], time-series modeling [5], general exponential smoothing [6]. These methods only can predict the linear load series and lack the ability to analyze the nonlinear character of load series. With the rapid development of artificial intelligence algorithm, the algorithms with strong self-learning ability, such as artificial neural network [7], BP neural network [8], simulated annealing algorithm [9], expert system, particle

* Corresponding author e-mail: yxwh_6726078@sina.com

swarm optimization [10], fuzzy inference [11], hierarchy matching [12], have been widely used in load forecasting. Though, artificial intelligence methods can deal with nonlinear relationship between the load and its relative factors, many of them still have some flaws. For example, expert system can not avoid the wrong expert knowledge, the shortcoming of ANN lies in over-fitting and long training time, fuzzy inference method still needs the expert' experience to generate. Recently, support vector machine (SVM), which is especially suitable for solving problems of small sample size, has also been applied for load forecasting [13]. Some improved SVMs have also been put forward to solve the concrete problems [14,15]. LS-SVM method can reduce the complexity of calculation effectively by converting quadratic program into linear equality sets, but the parameters of LS-SVM are mainly determined based on experience. Therefore, this paper will present a optimization algorithm to improve the parameters selection progress of LS-SVM.

Harmony search algorithm (HS) optimization algorithm [16,17], with the characteristic of overall, celerity and accuracy, has been successfully applied to solve problems such as performance optimization and reactive power optimization of power system. The HS optimization algorithm was introduced in this paper to solve hyper-parameters selection problem, and actual implementation result proved that the proposed algorithm can achieve higher prediction accuracy and better computational speed than methods in existing literature above. This paper is organized as follows. In section 2, GRNN network is used to choose the key factors that affect load forecasting. The LS-SVM is described in section 3. Section 4 uses HS optimization algorithm to quantitative identify the hyper-parameters of LS-SVM and to achieve its adjustment automatically. In section 5, through the empirical study, the accuracy and speed of load forecasting methods, including artificial neural networks (ANN), least squares support vector machine (LS-SVM), particle swarm optimization (PSO) and HS-LS-SVM, were analyzed comparatively. Finally, the conclusions are presented in section 6.

2 Select the Key Factors of Affecting the Short-Term Load Forecast Based on GRNN

When doing short-term load forecast, the factors that affect the electricity consumption should be considered carefully. According to previous studies, these possible influencing factors were summarized and divided into two categories here: economic factors and non-economic factors. The fluctuation of economic factors would affect the electricity consumption, even consumers' electricity consumption habits, such as salary(in month), power price, alternative energy price, power utility manners (electrical equipment and period), and non-economic factors also play an important role in affecting the electricity consumption, such as temperature, humidity, rainfall and holiday. Furthermore, electricity consumption is also influenced by demand side management measures, power price control, electric power sales promotion, national policy(especially the national industrial policy and energy policy implemented in the month), and so on.

General regression neural network (GRNN) proposed by Donald F. Specht in 1991 is a neural network architecture that can solve any function approximation problem [18]. If the training sample is provided, the network structure can be determined, and the connection weights of BP network can be obtained used by GRNN. Its learning rate is fast. It can approach a discontinuous function at any accuracy. The network converges to the optimized regression surface which is accumulated on the most samples. It also has better learning performance in the condition that irregular data are used in the network. Besides, there is only one artificial parameter, which maximize avoid influence of subjective assumption on the forecast results.

The network topological structure of GRNN consists of four layers: input layer, pattern layer, summation layer and output layer, as shown in Figure 2.1. A=[a_1,a_2,...,a_m]^T is the input vector, and B=[b_1,b_2,...b_l]^T is the out put vector.

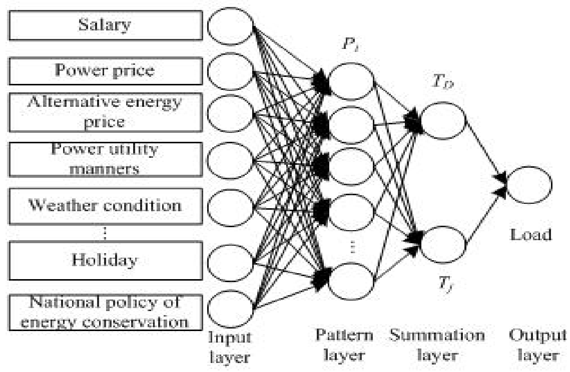


Fig. 2.1 The GRNN network topology of influence factors choosing in short term load forecasting

Based on GRNN, the steps of selecting the key factors of short-term load forecasting are described as follow:

It is assumed that the joint probability density function is $g(a, b)$ with variables a and b. The observed value of a is A. Then, relative to a, the regression of b can be expressed as:

$$\bar{B} = \frac{\int_{-\infty}^{+\infty} bg(A,b)db}{\int_{-\infty}^{+\infty} g(A,b)db} \tag{1}$$

The unknown probability density function $g(a, b)$ can be obtained through the observed value of a and b. The nonparametric estimation can be expressed as:

$$g^{\Lambda}(A,B) = \frac{1}{(2\pi)^{\frac{m+1}{2}}\sigma^{m+1}n}\times \sum_{i=1}^{n}\exp[-\frac{(A-A_i)^T(A-A_i)}{2\sigma^2}]\exp[-\frac{(B-B_i)^2}{2\sigma^2}] \quad (2)$$

Where A_i and B_i are the observed values of random variables a and b respectively. n denotes the number of samples; m denotes the dimensions of random variable a, and $g(a, b)$ is replaced by $g^{\Lambda}(A,B)$.

Based on Eq.(1), the estimation value $\overline{B}(A)$ can be obtained by exchanging the sequence of integration and summation:

$$\bar{B}(A) = \frac{\sum_{i=1}^{n}\exp[-\frac{(A-A_i)^T(A-A_i)}{2\sigma^2}]\int_{-\infty}^{+\infty} y\exp[-\frac{(B-B_i)^2}{2\sigma^2}]db}{\sum_{i=1}^{n}\exp[-\frac{(A-A_i)^T(A-A_i)}{2\sigma^2}]\int_{-\infty}^{+\infty}\exp[-\frac{(B-B_i)^2}{2\sigma^2}]db} \quad (3)$$

Where $\overline{B}(A)$ is the weighted mean of whole samples. The weight of observed value B_i is the exponent of the squared CityplaceEuclid distance between corresponding samples A_i and A.

Step 1: The sample data is set as input data. The number of input neurons is equal to the number of factors influencing load forecasting, and each neuron presents data to the second layer directly, namely the pattern layer.

Step 2: The number of pattern neurons is equal to the number of cases in the training set n. The typical pattern neuron i attains the data from the input neurons and computes an output P_i using the transfer function:

$$P_i = \exp[-\frac{(A-A_i)^T(A-A_i)}{2\sigma^2}], i = 1,2,\cdots,n \quad (4)$$

where σ denotes the smoothing parameter, and P_i denotes the weight of observed value B_i.

Step 3: The summation neurons include two kinds of neurons. One is simple arithmetic summation, which sums the outputs of pattern neurons. The weight between the pattern neurons and the summation is set at 1, and the transfer function T_D can be expressed as:

$$T_D = \sum_{i=1}^{n} P_i \quad (5)$$

The other is weighted summation. The weight between the neuron i of pattern neurons and the neuron j of summation neurons is the element j of output sample B_i.The transfer function of summation neurons can be expressed as:

$$T_j = \sum_{i=1}^{n} b_{ij}P_i, j = 1,2,\cdots,l \quad (6)$$

Step 4: The number of output neurons is equal to the dimension of output vector, l. Then, output neuron performs the following division to obtain the GRNN regression output b_j:

$$b_j = T_j/T_D, j = 1,2,\cdots,l \quad (7)$$

Step 5: The testing samples are chosen by extracting some continuous samples randomly from training samples. It is assumed that the smoothness factor is increased by $\Delta\sigma$ and the range is [σ_{min}, σ_{max}]. In the learning sample, the other samples are used for training except one sample. The error between estimation value and actual value is obtained. The process is repeated until each sample is excluded once, and the error series could be obtained. The MSE (mean square error, E) is applied to evaluate the network performance and expressed as Eq.(8). The optimized smoothness factor which has the minimum error is used for the final network training:

$$E = \frac{1}{n}\sum_{i=1}^{n}[\bar{B}_i(A_i) - B_i]^2 \quad (8)$$

The factor weight P_i influencing load forecasting is determined while the smoothness factor is optimized. If P_i meets the constraint of Eq.(9), the factor i can be chosen as one of the final influencing factors, which is used for load forecasting:

$$P_i \geq \frac{T_D}{n}, i = 1,2\cdots,n \quad (9)$$

3 Least Square-Support Vector Machine (LS-SVM)

The standard LS-SVM algorithm was introduced as follows [21]. Assume a set of training set is given like {$\boldsymbol{a}_i$, $\boldsymbol{b}_i$}(i=1,2,...,N), with the input $\boldsymbol{a}_i \in F^N$ and the output $\boldsymbol{b}_i \in F^N$. The following regression model is constructed by using nonlinear mapping function $\boldsymbol{f}$(.), which maps the input data to higher dimensional feature space.

$$g(a_i)=H^{\mathrm{T}}f(a_i)+m \quad (10)$$

Where $\boldsymbol{H}$ and $\boldsymbol{m}$ are parameters needed to be determined, which can be calculated by minimizing the following function:

$$F=0.5||H||^2+\zeta F_1 \quad (11)$$

Where ζ is a regularization parameter and determines the trade off between minimizing the training error and minimizing model complexity; F_1 is the error term, i.e. empirical risk in learning theory. The optimization problem of the LS-SVM regression can be expressed as:

$$\begin{cases} \min \ J[H,e_i] = 0.5\|H\|^2 + 0.5\zeta\sum_{i=1}^{N} e_i^2 \\ s.t. \ \ b_i = H^T f(a_i) + m + e_i, i = 1,2,\cdots,N \end{cases} \quad (12)$$

And then, Lagrange function is adopted to solve this optimization problem.

$$L(H,\lambda_i,m,e_i)=J+\sum_{i=1}^{N}\lambda_i[b_i-H^Tf(a_i)-m-e_i] \quad (13)$$

Where λ_i is Lagrange multipliers, $\lambda_i \geq 0$; $\boldsymbol{e}_i$ is training errors. According to Karush–Kuhn–Tucker conditions, the first order partial derivatives of Lagrange function can be expressed as:

$$\begin{cases} \frac{\partial L}{\partial H}=0 \\ \frac{\partial L}{\partial \lambda_i}=0 \\ \frac{\partial L}{\partial m}=0 \\ \frac{\partial L}{\partial e_i}=0 \end{cases} \quad (14)$$

And then, Eq.(15) and Eq.(16) can be obtained:

$$H=\sum_{i=1}^{N}\lambda_i f(a_i), \sum_{i=1}^{N}\lambda_i=0, \lambda=\zeta e_i \quad (15)$$

$$b_i-H^{\mathrm{T}}f(a_i)-m-e_i=\mathbf{0} \quad (16)$$

When the variable $\boldsymbol{H}$ and $\boldsymbol{e}_i$are removed, the optimization problems can be described as a linear system.

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & k(a_1,a_1)+\frac{1}{r} & \cdots & k(a_1,a_N) \\ & & \vdots & \\ 1 & k(a_N,a_1) & \cdots & k(a_N,a_N)+\frac{1}{r} \end{bmatrix} \begin{bmatrix} m \\ \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} 0 \\ b_1 \\ \vdots \\ b_N \end{bmatrix} \quad (17)$$

λ_i and $\boldsymbol{m}$ can be obtained by solving Eq.(17). The LS-SVM regression model can be expressed as:

$$\hat{b}(a)=\sum_{i=1}^{N}\lambda_i\varphi(a_i,a)+m \quad (18)$$

Where $\varphi(\boldsymbol{a}_i,\boldsymbol{a})=\boldsymbol{f}(\boldsymbol{a}_i)^T\boldsymbol{f}(\boldsymbol{a})$ is the kernel function, which follow Mercer's theory. The common examples of kernel function contain: 1) linear function, $\varphi(\boldsymbol{a}, \boldsymbol{a}_k)=\boldsymbol{a}_k{}^T\boldsymbol{a}$; 2) polynomial kernel function, $\varphi(\boldsymbol{a},\boldsymbol{a}_k)=(\boldsymbol{a}_k{}^T\boldsymbol{a}+1)^d$; 3) radial basis function (RBF) kernel function, $\varphi(\boldsymbol{a},\boldsymbol{a}_k)=\exp[-||\boldsymbol{a}-\boldsymbol{a}_k||^2/(2\mu^2)]$; 4) multi-layer perceptron (MLP) kernel function, $\varphi(\boldsymbol{a},\boldsymbol{a}_k)=\tanh(k\boldsymbol{a}_k{}^T\boldsymbol{a}+\eta)$.

According to the problems of training LS-SVM, proper parameter setting plays a crucial role in building a good LS-SVM regression model with high prediction accuracy and stability, such as regularization parameter ζ and the RBF kernel function parameter μ. In this research, these parameters can be called directly in the training phase through harmony search algorithm (HS) algorithm.

4 Short-Term Load Forecasting Based on HS-LS-SVM Optimization Algorithm

4.1. HS Algorithm. The harmony search algorithm (HS) is a meta-heuristic algorithm inspired by the improvisation process of music players[22]. The musicians in the band or orchestra are represented by the components of the solution vector. Perfect harmony occurs when each musician plays the perfect note. In the same way the perfect solution vector is found when the value of each component is optimal. A musician improvises new tones and tests them for harmony with the rest of the band. If the new improvisation works well the improvised tone is remembered for future use, otherwise the musician forgets the tone and tries a different improvisation. The harmony search algorithm mimics this behavior by keeping a matrix of the best solution vectors called the Harmony Memory (HM). The number of vectors that can be simultaneously remembered in the memory is known as the Harmony Memory Size (HMS) and is one of the algorithm's parameters that is set during initialization. The memory is organized as a matrix with each row representing a solution vector and the final column representing the vector's fitness. The steps of HS algorithm are described as follows:

Step (1). Initialize the optimization problem and algorithm parameters.

Step (2) Initialize the harmony memory (HM).

Step (3). Improvise a new harmony from the HM.

Step (4). Improvise a new harmony from best harmony (local search).

Step (5). Update the HM.

Step (6). Repeat Steps 3 ,4 , 5 until the termination criterion is satisfied.

4.2. The Hyper-Parameters Selection Based on HS. It should be noted that there are two key factors in the selection of hyper-parameters based on HS: 1) how to describe the hyper-parameters with HM; which will directly affect the searching efficiency and convergence speed; 2) how to define the fitness function.

In the proposed algorithm, HM is required to provide one potential solution, which is called hyper-parameters combination. Here, a hyper-parameters combination is defined as one dimensional combined vector x. For example, RBF: $\boldsymbol{x}=(\zeta, \mu)$. In this research, set $\boldsymbol{x}=(\log_2\zeta, \log_2\mu)$ for the reason that this method has a higher searching efficiency and has got more stable optimization result.

In this problem, through calculating the fitness value of HM, the fitness function can guide HS algorithm to move to better solution, which shows that it should be defined for different problems. The higher the fitness value is, the better the location is. The fitness function is defined as follow:

$$G=-\frac{1}{N}\sum_{j=1}^{N}\frac{\left|c_j-c_j'\right|}{c_j}\times 100 \quad (19)$$

Where c'_j and c_j denote training results and actual results respectively.

4.3. Modeling Progress. Before modeling, it needs to select the load sample data, and preprocess the data with the following method. Then, the sample data are normalized to [0, 1], which avoid the large amount of computation.

$$x'_{ij} = \frac{x_{ij} - x_{ij}^{\min}}{x_{ij}^{\max} - x_{ij}^{\min}} \tag{20}$$

Where $x'_{ij}$is the non-dimensional value of sample i; x_{ij} is the initial value; ${x_j}^{min}$ is the minimum value of property j of all the samples; ${x_j}^{max}$ is the maximum value of property j of all the samples.

Considering the better recognition and attainable of RBF kernel function, the RBF kernel function is selected to use in this study. The parameters of the HS algorithm are: the HMS is 20, the maximum iteration number is 1000, the dimension is 2, the trace back step number is 5, and the initial HM is distributed within [-5, 15] randomly.

Meanwhile, since the HS algorithm uses a kind of stochastic search technique, the optimization progress can ran one hundred iterations by calculating the training samples. In each iteration, the two-dimensional position of particle with the biggest value of fitness function is given to the optimal hyper-parameters combination noted (ζ,μ^2).

Set the normalized samples as the input data of LS-SVM model, then, forecast the load. Finally, output the forecasting result. The flow chart of LS-SVM hyper-parameters selection algorithm based on HS is shown in Figure. 4.1.

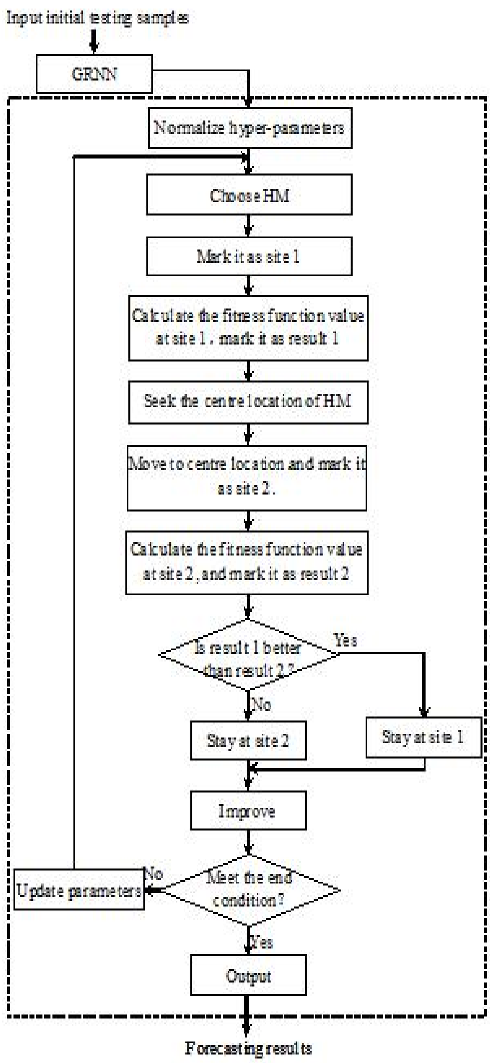


Fig. 4.1 Flow chart of LS-SVM hyper-parameters selection algorithm based on HS

5 Numerical Example

Data are chosen from power load database of a certain area in China. The power load datas from 1/12/2010 to 29/12/2010 are selected as training sample and used to verify the effectiveness and the advancement of the proposed method. Based on the Matlab 7.0 platform and simulate analysis of GRNN, the influence of factors on the error of training results are analyzed. Combining the error analysis and the weights obtained from GRNN, the factors which influence the error of training results much are chosen as the key factors, including ahead-day load data at period i, ahead-day load forecasting data at period i, weather data of sample day, weather data of forecasting day, holiday period, which are also the input variables. Hourly load of forecasting day is set as output variables. Input variables can be described as follow:

(1) $A=\{a_{1,1},a_{1,2},...,a_{1,24},a_{2,1},...a_{2,24},...a_{n,i}\}$ denotes the ahead-day load data of sample n at period i;

(2) $B=\{b_1,b_2,...,b_i\}$ denotes the ahead-day load forecasting data at period i;

(3) $C=\{c_{1,j},c_{2,j},\ldots c_{n,j}\}$ denotes the weather data of sample day n, j=1,2,3,4. $c_{n,1}$ denotes the low temperature of sample day n. $c_{n,2}$ denotes the high temperature of sample day n. $c_{n,3}$ denotes the average temperature of sample day n. $c_{n,4}$ denotes the rainfall of sample day n.

(4) $D=\{d_1,d_2,d_3,d_4\}$ denotes the weather data of forecasting day, including low temperature, high temperature, average temperature, rainfall.

(5) $E=\{0,1\}$. E=0 denotes weekdays (Monday to Friday); E=1 denotes weekends (Saturday and Sunday).

There are 696 samples in 29 days, of which 672 samples from 1/12/2010 to 28/12/2010 are set as training data. Meanwhile, 24 forecasting results on 29/12/2010 are obtained to reflect different load.

The optimal hyper-parameter combination (ζ,μ^2)=(88.3,131.7) can be got by using the proposed method in this paper. And the forecasting results have better forecast precision and mean absolute percentage

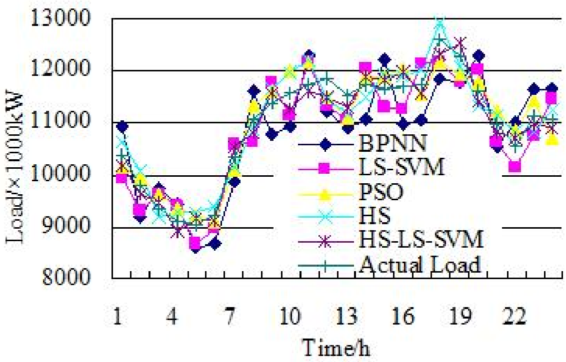


Fig. 4.2 Fitting curves of five methods

error (MAPE) with the error is only 1.78% ($\frac{1}{72}\sum_{i=1}^{72}\left(\left|b_i-\hat{b}_i\right|\right)/b_i = 1.78\%$, where $\hat{b}_i$ and $b_i$denote the testing output value and actual value respectively).

To assess the reasonable of the proposed method, the comparative analysis of the results from three methods (BPNN, LS-SVM, PSO and HS-LS-SVM) is made. Adopt the "8-17-1" network of Levenberg-Marquardt (LM), and choose Tansig function $\left(f(x)=2/\left(1+\mathrm{e}^{-2x}\right)-1\right)$ and Purelin function $(f(x)=\mathrm{x})$ as the transfer functions of pattern layer and output layer respectively. The training parameters are set as follows: the learning rate is 0.01, the target error is 0.001, the maximum iteration number is 2000. Comparing with the HS hyper-parameter selection algorithm, the network searching algorithm applies the same parameter, but its hyper-parameters search range is in the scope of [-5, 15] with the step is 1. The fitting curve of load forecasting results on 29/12/2010 and the relative error distributions of different algorithms are shown in Table 1 and Figure 4.2. The training time in table 1 is the average training time after operation 100 times.

As shown in Table 1, BPNN algorithm has the highest MAPE, up to 5.13%. The MAPE of HS-LS-SVM is the lowest, only 1.76%. The MAPE of LS-SVM and PSO is respectively 3.55% and 2.53% which are between the above two values. The training time of HS-LS-SVM is 31.79s, and the training time of LS-SVM and PSO is respectively 144.68s and 82.36s. Based on results above, the analysis are as follows: (1) The MAPE of HS-LS-SVM is 3.37 percentage points lower than BPNN. It shows that LS-SVM has obvious advantage in solving small sample set regression problem comparing to the BPNN algorithm. That is because LS-SVM algorithm can satisfy the SRM(Structural Risk Minimization) principle proposed by Vladimir Vapnik and Alexey Chervonenkis in 1974, has good generalization ability and overcomes the deficiencies of ANN algorithm which has high requirement for training sample number and quality. (2) The MAPE of HS-LS-SVM is 1.79 percentage points lower than LS-SVM. And the training time of HS-LS-SVM is only 21.97% of LS-SVM, i.e. the training speed is 4.55 times higher than the latter. So, it suggests that in both the solution quality and training speed HS-LS-SVM algorithm has much higher efficiency than LS-SVM. (3) The MAPE of HS-LS-SVM is 0.77 percentage lower than PSO algorithm. And the training time of HS-LS-SVM is only 38.60% of PSO, i.e. the training speed is 2.59 times higher than the latter. So, it suggests that compared with traditional advanced short-time load forecasting algorithm, a significant advantage of HS-LS-SVM algorithm is the better computational speed which is more practical for short term load forecasting. (4) Compared with LS-SVM, introduction of HS makes error rate dropped by 50.42%, while introduction of HS makes training time dropped by 78.03%. It indicates that LS-SVM optimized by HS achieve higher prediction accuracy and better computational speed. Compared with the prediction accuracy improvement, the computational speed improvement is more significant. Unlike other optimal algorithms, the way to choose search scale for HM in HS algorithm is based on probability distribution, which can effectively breakthrough the limitation of local extremum. At the same time, using the interactive mode of HM can not only learn from their past way, but also rapidly approach the optimal value by borrowing the experience from other HM. So, the example study shows that comparing with other optimization algorithms, the LS-SVM optimization algorithm based on HS has obviously advantages in both the calculation accuracy and calculation speed.

Conclusions

Considering the economic factors and non-economic factors which affect the short-term load forecasting, the key factors are selected by GRNN. Then, the proposed method–HS-LS-SVM is used to forecast load. From the results of experiment, we can get the following conclusions: (1) Through analysis of load influencing factors with GRNN, the key influencing factors are selected as the output variable, so the accuracy of forecasting model can be improved. (2) According to the problems of LS-SVM parameters, the HS algorithm is applied to optimize these parameters. The experiments results show that the proposed method can automatically extract parameters which have high recognition rate and fast convergence speed. (3) A new short-term load forecasting approach based on HS-LS-SVM is proposed in this paper. The experiments results show that the proposed method achieves higher precisions and faster speed than BPNN, LS-SVM and PSO, and its correctness and effectiveness are also verified. As a heuristic hybrid algorithm, the proposed method has a bright application future, and can be applied not only in short-term load forecasting, but also in forecasting of other areas.

Table 5.1 Comparison of 24-hour forecasting results among different methods of December 29,2010

	BPNN		LS-SVM		PSO		HS		HS-LS-SVM		Actual load
Time	Fore-casting results ($\times10^3$ kW)	Rela-tive error (%)	Fore-casting results ($\times10^3$ kW)	Rela-tive error (%)	Fore-casting results ($\times10^3$ kW)	Rela-tive error (%)	Fore-casting results ($\times10^3$ kW)	Rela-tive error (%)	Fore-casting results ($\times10^3$ kW)	Rela-tive error (%)	($\times10^3$ kW)
1:00	10925.97	5.27	9939.97	4.23	10151.7	2.19	10608.36	2.21	10169.34	2.02	10379
2:00	9198.99	5.96	9294.86	4.98	9942.42	1.64	10081.33	3.06	9620.6	1.65	9782
3:00	9741.51	4.31	9633.18	3.15	9624.77	3.06	9191.44	1.58	9478.15	1.49	9339
4:00	9425.28	3.62	9400.72	3.35	9344.32	2.73	9287.93	2.11	8931.36	1.81	9096
5:00	8584.34	5.03	8687.38	3.89	9242.38	2.25	9252.32	2.36	9164.64	1.39	9039
6:00	8674.58	6.17	8951.93	3.17	9130.36	1.24	9405.86	1.74	9110.95	1.45	9245
7:00	9868.66	4.54	10563.37	2.18	10118.83	2.12	10126.07	2.05	10539.59	1.95	10338
8:00	11620.18	4.97	10623.88	4.03	11324.61	2.3	10839.74	2.08	10835.32	2.12	11070
9:00	10790.5	5.28	11754.27	3.18	11605.03	1.87	11667.69	2.42	11554.91	1.43	11392
10:00	10917.15	5.61	11135.74	3.72	12014.76	3.88	11978.91	3.57	11267.6	2.58	11566
11:00	12291.37	4.75	12154.08	3.58	12166.98	3.69	12215.09	4.10	11602.58	1.12	11734
12:00	11226.93	5.25	11338.31	4.31	11499.45	2.95	11450.87	3.36	11486.42	3.06	11849
13:00	10904.75	5.39	11052.28	4.11	11098.39	3.71	11194.05	2.88	11311.62	1.86	11526
14:00	11066.97	5.58	12064.43	2.93	11897.99	1.51	11495.96	1.92	11906.19	1.58	11721
15:00	12221.89	4.99	11284.79	3.06	11881.97	2.07	11939.01	2.56	11798.15	1.35	11641
16:00	10967.27	6.07	11265	3.52	12043.79	3.15	11886.17	1.80	11987.75	2.67	11676
17:00	11070.13	5.73	12129.34	3.29	11536.32	1.76	11988.43	2.09	11584.47	1.35	11743
18:00	11856.33	5.79	12159.63	3.38	12184.8	3.18	12918.50	2.65	12343.37	1.92	12585
19:00	11769.87	4.31	11789.55	4.15	11940.84	2.92	12072.45	1.85	12537.39	1.93	12300
20:00	12284.2	5.88	11998.79	3.42	11770.23	1.45	11327.03	2.37	11431.45	1.47	11602
21:00	10528.18	4.01	10627.99	3.1	11196.13	2.08	11146.78	1.63	10799.09	1.54	10968
22:00	11022.85	4.76	10155.83	3.48	10847.14	3.09	10820.82	2.84	10698.77	1.68	10522
23:00	11641.51	4.53	10753.89	3.44	11420.87	2.55	10812.91	2.91	10973.29	1.47	11137
24:00	11646.23	5.31	11453.81	3.57	10700.62	3.24	11255.85	1.78	10901.96	1.42	11059
Average relative error(%)	5.13		3.55		2.53		2.41		1.76		——
Training time	——		144.68		82.36		94.03		31.79		——

6 Acknowledgement

The work described in this paper was supported by National Science Foundation of China (NSFC)(71271082), The National Soft Science Research Program (2012GXS4B064) and Energy Foundation of U.S (G-1006-12630).

References

[1] Zhang W.L., Tang G.F., ZHA K.P., .He Z.Y, *Application of Advanced Power Electronics in Smart Grid*, *Proceedings of the CSEE*. Mag. **30**, 1-7 (2010).

[2] Zeng Ming, Xue Song, CityplaceZhu Xiaoli, *country-regionChina's 12th Five-Year Plan Pushes Power Industry in New Directions*, *Power. Mag*.**156**, 50-55 (2012).

[3] Amjady, N. Short-term hourly load forecasting using time-series modeling with peak load estimation capabilit, IEEE Transactions on Power Systems. Mag. 16, 798-805 (2002).

[4] Papalexopoulos, A.D., Hesterberg, T.C, *A regression-based approach to short-term system load forecasting*, *IEEE Transactions on Power Systems. Mag.* **5**, 1535-1547 (1990).

[5] Chiu, C.C., Kao, L.J., Cook, D.F. Combining a neural network with a rule-based expert system approach for short-term power load forecasting in Taiwan. Expert Systems with Application., Mag. 13, 299-305 (1997).

[6] Christiaanse, W.R, *Short-term load forecasting using general exponential smoothing*, *IEEE Transactions on Power Apparatus and Systems*. *Mag*. **PAS-90**, 900-911 (2007).

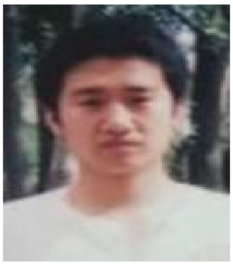
[7] Bao Jian, Zhou LiangJie, Yan Yi, *Analysis on complexity of neural networks using integer weights*. *Applied mathematics & Information scineces Mag.* **6**, 317-323 (2012).

[8] Xiao, Z., Ye, S.J., Zhong, B., Sun, C.X, *BP neural network with rough set for short term load forecasting*, *Expert Systems with Applications. Mag*. **36**, 273-279 (2009).

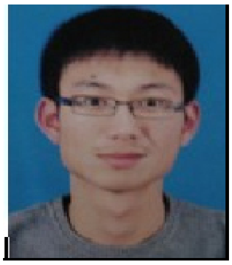
[9] Liao, G.C., Tsao, T.P, *Application of a fuzzy neural network combined with a chaos genetic algorithm and simulated annealing to short-term load forecasting*, *IEEE Transactions on Evolutionary Computation. Mag.* **10**, 330-340 (2006).

[10] Wang Yanming, Wang Deming, Zhong Xiaoxing, *Modified Particle Swarm Algorithm for Radiation Properties of Semi-transparent Rectangular Material*, *Applied mathematics & Information scineces*. *Mag*. **5**, 227-233 (2011).
[11] Mamlook, R., Badran, O., Abdulhadi, E. A fuzzy inference model for short-term load forecasting. Energy Policy. Mag. 37, 1239-1248 (2009).
[12] Wang, Z.Y., Cao, Y.J, *Case-based reasoning for short-term load forecasting based on hierarchy matching*. *Journal of Zhejiang University (Engineering Science)*. *Mag*. **41**, 1598-1603 (2007).
[13] Wu, Q, *Hybrid model based on wavelet support vector machine and modified genetic algorithm penalizing Gaussian noises for power load forecasts*. *Expert Systems with Applications*. *Mag*. **38**, 379-385 (2011).
[14] Wu, Q., Law, R, *The forecasting model based on fuzzy novel* ν*-support vector machine*, *Expert Systems with Applications*. *Mag*. **38**, 12028-12034 (2011).
[15] Ismail, S., Shabri, A., Samsudin, R, *A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting*, *Expert Systems with Applications*. *Mag*. **38**, 10574-10578 (2011).
[16] Khazali A.H., Kalantar M, *Optimal reactive power dispatch based on harmony search algorithm*, *International Journal of Electrical Power & Energy Systems*. *Mag*. **33**, 684-689 (2011).
[17] Khorram E, Jaberipour M, *Harmony search algorithm for solving combined heat and power economic dispatch problems*, *Energy Conversion and Management*. *Mag*. **52**, 1550-1554 (2011).
[18] Polat O, Yildinm T, *FPGA implementation of a General Regression Neural Network:An embedded pattern classification system*, *Digital Signal Processing*. *Mag*. **20**, 881-886 (2010).
[19] Chen, Y.H., Cao, G.Y., Zhu, X.J. *LS-SVM model based nonlinear predictive control for MCFC system*, *Journal of Zhejiang University: Science A*. *Mag*. **9**, 748-754 (2007).
[20] Arasimham S.V.L., Ramalinga Raju M., Srinivasa Rao A, *Optimal Network Reconfiguration of Large-Scale Distribution System Using Harmony Search Algorithm*, *IEEE Transactions on Power Systems*. *Mag*. **26**, 1080-1088 (2011).

Ming Zeng received his Electrical Technology Power Station Engineering Master degree from North China Electric Power University, 1984. He is now a Professor at the same University. The project DSM mechanism Aimed at Energy Conservation and Emissions Reduction in China he charged in won the first prize of National Energy Administration Soft Science Outstanding Research Achievements. His research interests include energy and power system economics, low-carbon electricity, power system planning, power market, power system reliability and power system investment.



Song Xue received his bachelor degree in School of Economics and Management in North China Electric Power University, China, 2009. He is currently pursuing Ph.D degree at School of Economics and Management in the same University. His research interests include low-carbon electricity, power system economics and optimization, power market and power system coordination expansion planning.



Zhijie Wang received his bachelor degree in School of Economics and Management in North China Electric Power University, China, 2011. She is currently pursuing master degree at School of Economics and Management in the same University. His research interests include power system economics and optimization, power market and Power company decision and optimization. And he is the corresponding author.

Xiaoli Zhu received bachelor degree in School of Economics and Management in North China Electric Power University, China, 2010. She is currently pursuing master degree at School of Economics and Management in the same University. Her research interests include power system economics and optimization, power market and low-carbon electricity.

Ge Zhang received her Ph.D degree in Management from School of Economics and Management, North China Electric Power University in 2012. She is currently an associate professor in North China Electric Power University. Her research interests are in the areas of power enterprise strategic cost management and electric power enterprise performance evaluation.