The Application of Joint Optimization Method using GA for Load Forecasting

Junsong Qin[1, a], Dongxiao Niu[1, b], Jinpeng Qiu[1,c] and Ling Ji[2, d]

[1]School of Economics and Management, North China Electric Power University, Beijing 102206, China;

[2]School of Economics and Management, Beijing University of Technology, Beijing 100124, China.

[a]285297091@qq.com, [b]niudx@126.com, [c]1550872694@qq.com, [d]hdjiling@126.com

Keywords: Least square support vector machine (LSSVM); genetic algorithm (GA); parameter optimization; phase space; short-term load forecasting.

Abstract. Since load forecasting plays an important role in the planning and operation of power industry, substantial efforts are made in improving the accuracy and reliability of load forecasting. In this paper, we develop a novel hybrid approach based on phase space reconstruction and least square support vector for the short-term load forecasting. However, the proper parameters in phase space reconstruction and least square vector machine have a significant effect on the forecasting performance, and there is no standard solution for the parameter estimation problem. Therefore, in this paper, the genetic algorithm (GA) approach is employed to optimize the parameters of both phase space reconstruction and least square support vector machine together. The experimental results suggest that the joint optimization parameter is superior to the separate optimization solutions.

Introduction

Electric Power Demand forecasts are a key issue in the electricity industry, as they provide the basis for making decisions in power system planning and operation. It is therefore necessary to develop new techniques for reducing the uncertainty of the predictions [1]. For the short-term load forecasting, the methods can be classified into three types: (*i*) trend method, which is non-causal model, and does not explicitly explain how the projected variable is determined; (*ii*) time series method, which get the prediction from the historical patterns of the data; (*iii*) neural network techniques, a popular method, are able to capture and represent complex the input-output relationships, especially for non-linear issues [2-3]. The advantage is the ability to learn these relationships directly from the data being modeled.

Load series is a typical chaotic time series. Methods such as whole domain method, local region method, neural network and support vector machine are commonly applied for chaotic time series forecasting. Among them, least square support vector machines (LSSVM), a machine learning algorithm based on the structural risk minimization principle, can address the issues including nonlinear, high dimension and the local minimum and have strong generalization ability, which over comes the fitting defects in neural network. Therefore, it has achieved very good effect in chaos time series prediction research. However, the learning and generalization ability are depend on the parameters. In fact, due to lacking of uniform standard parameter selection and theoretical guidance, the performance of different algorithm is various. The selection of LSSVM parameters is still an unsolved problem. In this paper, parameter selection exists in both phase space reconstruction and LSSVM, which determine the precision of chaos time series, so joint optimization should be considered. At currently, most parameters optimization is separated, implying that the correlation among parameters is ignored. According to the defects, the forecasting model based on the joint optimization method GA-LSSVM is proposed.

The rest part of this paper is organized as following: in Section 2, the basic concepts of phase space reconstruction and LSSVM are briefly reviewed. The joint parameter optimization via genetic algorithm is designed in Section 3. In Section 4, the proposed model is applied to short-time load

forecasting and its performance is compared with that of C-C-LSSVM. Conclusions will be the last section of the paper.

Methodology

Phase Space Reconstruction. Though phase space reconstruction, the time series with chaos characteristic is turned into low order nonlinear dynamic system, so that it approximately recovers the system's chaos attractor [4]. We can Phase space is reconstructed according to the delay coordinate method proposed by Takens and Packard. We can mine the information from the chaos time series data by determining the optimal time delay (τ) and embedding dimension (m) to restore the impulsion system in chaotic time series. Then, the reconstructed data is utilized for learning and modeling to forecast the future value. So the reconstruction quality impacts the forecasting model and results directly [5-6].

The observed chaos time series is defined as: $\{x(t)\}$, $t = 1, 2, ..., n$, n is the number of samples. According to Takens theory, the chaos time series can be reconstructed via appropriate τ and m:

$$X_n = \left(x_n, x_n + \tau, ..., x_n + (m-1)\tau \right) \tag{1}$$

From Eq.(1), it's obvious that the reconstruction results depend on the parameters τ and m. According to Takens theory, τ can be arbitrary, if the chaotic time series is infinite and noise free. That is, the value of τ has nothing to do with m. Whereas, such chaotic time series is almost inexistent in practice, especially to short-term load. Hence, the selection of τ is important to the phase space reconstruction, and both τ and m are correlative. If the embedding dimension (m) is too large, the space will become much more complex; while, it is too small, the original system dynamics characteristics can't be expressed well [7-8].

There are various methods for embedding dimensions estimation, such as false neighbors method, Cao method, saturated correlation dimension method, and so on. Paper [9] employed Cao method to recognize embedding dimension and the mutual information method is used to recognize time delay, and its application on the gas emission rate is feasible. Paper [10] applied the so-called false nearest-neighbor method for a multivariate local polynomial prediction model. But the performance is various due to the different data, and the judgment standard is subjective. It is clear that there is a lack of optimal selection method with strong popularity.

Theory of Least Square Support Vector Machines. The Support Vector Machine (SVM), a machine learning algorithm based on statistical learning theory, is proposed by V. Vapnik in 1990s, and developed rapidly. The basic idea of SVM is to map the input space data into a higher dimensional feature space by nonlinear mapping, and the practical problem is converted into quadratic programming problem with the inequality constraints. Since SVM can solve problems including the small sample, nonlinear, high dimension and the local minimum, it has been widely applied in pattern classification, regression forecast, probability estimate and control theory fields [11-12].

However, the complexity of SVM has nothing to do with the dimension of the input vector, it actually depend on the number of the samples. More the sample is, more complex the corresponding solution process for the quadratic programming problem, which also leads to slower calculation speed. This limits the application range of the support vector machine. Later, Suykens *et al* [13] proposed the least square support vector machine (LSSVM) based on the standard SVM. In that way, the loss function is set to the error square, besides the inequality constraints turned into equality constraints to reduce the parameters to confirm, and the solution of quadratic programming is transformed into solving the linear KKT (Karush-Kuhn-Tucker) equations, which reduces the complexity of the solution and broaden its application [14-16].

For training sample set $\{(x_i, y_i)\}$, ($i = 1, 2, ..., n$), x_i and y_i are the input and output sample respectively, $x_i \in R^n$, $y_i \in R$. Though nonlinear mapping function $\varphi(\cdot)$, the samples are mapped into high dimensional feature space, and the optimal linear regression function is described as:

$$f(x) = w^T \varphi(x) + b \tag{2}$$

where, w is the weight vector of feature vector, and b is the deviation.

Regarding the structural risk minimization principle, the LSSVM regression function of Eq.(2) is:

$$\min J(w, \xi) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{m} \zeta_i^2 \tag{3}$$

$$s.t. \ y_i = w\varphi(x) + b + \zeta_i, \ i = 1, 2, ..., n$$

where, γ is punishment factor to balance the training error and the complexity of the model, ζ_i is the slack variable.

Introducing Lagrange multiplier, the constrained optimization problem above changes into unconstrained dual space optimization problem, that is:

$$L(w, b, \zeta, \alpha) = \min \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^{n} \zeta_i^2 + \sum_{i=1}^{n} \alpha_i \left(w^T \varphi(x) + b + \zeta_i - y_i \right) \tag{4}$$

where, α_i are the Lagrange multipliers.

According to KKT condition, we can obtain:

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i \varphi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i = 0 \\ \dfrac{\partial L}{\partial \zeta_i} = 0 \rightarrow \alpha_i = \gamma \zeta_i \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + \zeta_i - y_i = 0 \end{cases} \tag{5}$$

Eliminating w and ζ_i, we can obtain the following matrix solution:

$$\begin{bmatrix} 0 & Q^T \\ Q & K + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ A \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{6}$$

where, $Q = [1, ..., 1]^T$, $A = [a_1, a_2, ..., a_n]^T$, $Y = [y_1, y_2, ..., y_n]^T$.

Under the Mercer conditions, the kernel function is defined as:

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \tag{7}$$

Although the function satisfying Mercer condition can be used as kernel function, the performance by different kernel function is various. Here, the radial base kernel function will be the best choice:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \tag{8}$$

where, σ denotes the width of the kernel function.

It can be seen from above, the learning ability of LSSVM based on radial base kernel function is mainly determined by γ and σ.

Joint optimal design of parameters

The joint optimization of τ, m, γ and σ is obviously a multi-parameter combinatorial optimization problem. If we use enumeration methodto to solve it, the calculation will be huge sometimes even can't be achieved. To reduce the complexity in parameter optimization, traditionally, we fill find each parameter's optimal level, then transform the multi-parameter problem into single-parameter optimization problem. While in this way, the interaction between parameters will be ignored. As a result, the combination of each parameter's best level may not the global optimal parameter.

The GA is a stochastic optimization algorithm originally motivated by the mechanisms of natural selection and evolution of genetics [17].Though natural selection, crossover and mutation mechanism, it can realize the population evolution, and can search the global optimal solution in relatively short time, due to its implied parallelism and powerful global search ability. Regarding its advantages, we adopt GA to find the best combination of τ, m, γ and σ.We now describe the details on employing the GA to find the optimal parameters.

Encoding. Binary system is adopted here for individual encoding. Since the radius base kernel function is selected as LSSVM kernel function, the individual in GA is consist of τ, m, γ and σ four parts.

The binary string of individual can be transformed into actual parameter expressed in decimal system though the following:

$$p = \min(p) + \frac{\max(p) - \min(p)}{2^l - 1} \times d \tag{10}$$

where, p is the decimal value of parameter; $\min(p)$ and $\max(p)$ mean the minimum and maximum value of parameter respectively. l is the length of the binary string; d is the decimal value of binary string.

Population Initialization. The initial population of basic genetic algorithm is randomly generated, and the distribution space of this initial population is always uncertain and uneven, which may not include the information of global optimal solution very well and leads to early convergence and local optimal problem easily. To avoid these problems, here we create the initial population by uniformly design to ensure the initial population divers and distributed uniformly [18].

Fitness Function. The search object of GA is to find the most suitable τ, m, γ and σ to improve the forecasting accuracy of chaotic time series and reduce the forecast error. Therefore, the fitness function can be related to the prediction ability. Let the forecasting mean squared error (e_{MS}) from the parameters in the jth group written as:

$$e_{MS} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{11}$$

where, n denotes the number of testing sets in chaotic time series; y_i and $\hat{y}_i$ represent the actual value and forecasting value respectively.

The fitness function of the jth individual can be defined as:

$$f(j) = \frac{1}{e_{MS}(j)} \tag{12}$$

Genetic Operator Design. (1) Selection.The selection strategy is based on the individual's fitness value. Sort the individual by its fitness value. According to the individual keep strategy, R individuals

with better fitness will enter to the next generation, and the rest (n-R) will reproduce the new individual after crossover and mutate for the next generation.(2) Crossover. This is done by the exchange between two mating parents. Random places are selected on each part and the strings between these places are exchanged. After this, the new generated individual need to be tested to find if the value of parameters is within the reasonable range, if not, crossover operation should be done again.The crossover operation is applied with the probabilistic values (the cross rate) from 0.2 to 0.8 [19]. (3) Mutation. In order to avoid a local optimum the mutation operation is served as a tool to introduce variation into a population by applying small changes to the solution [20]. Here, we use the traditional inversion mutation as mutation operation. Then test whether the new individual after mutation within the variable range, if not the mutation operation need be repeated.

Joint Optimization of Parameters. First, let the initial population of GA as the parameter of the forecasting model, including τ, m, γ and σ. Use them for the phase space reconstruction and LSSVM training and prediction, and we can obtain a series of testing error e_{MS}. Then calculate the fitness value of each individual, and generate the next generation by selection, crossover and mutate operation. The process is repeated until the stop criterion is satisfied. Finally, use the optimal parameter to build the forecasting model. The detail process can be divided into the following steps.

Step 1: Determine the initial feature space composed by all the parameters, and set the evolution iteration t = 0;

Step 2: Encode the parameters including τ, m, γ and σ;

Step3: According to uniformly design, generate the initial population with m individuals;

Step 4: transform the individual's binary string into actual parameter expressed in decimal system. The parameters τ and m are used to reconstruct the chaotic time series, then LSSVM is trained with γ and σ. Record the forecasting precision under each group parameters, and calculate the fitness value of each individual;

Step 5:keep the best R individuals with good performance into the next generation, and conduct the selection, crossover and mutation operation for the rest individual to generate the new population, and $t = t + 1$;

Step 6: Judge whether the stop criterion is satisfied. If it does, continue the process to step (7); otherwise, back to step (3), repeat the process above;

Step 7: Obtain the parameters of τ, m, γ and σ from the optimal individual. Conduct the phase space reconstruction of the chaotic time series with the optimal τ and m. The optimal γ and σ are used as the parameters of LSSVM to train and forecast the reconstructed series.

Case study and results analysis

Experiment is carried out here to test the effectiveness of the proposed model. The hourly load data ranging from 2010-03-15 to 2010-04-28 is collected from a grid company. We use them to forecast the 24 hour load on April 29th, 2010.The main parameters of genetic algorithm are set as following: radial basis kernel function is for LSSVM; the initial population is 70; the iteration is set as 1000, the cross rate and mutate rate are 0.8 and 0.04 respectively. The punishment factor $\gamma \in [1,1000]$, and $\sigma \in [0.1,10]$.

In order to verify the power of this new joint parameter optimization, we employ C-C-LSSVM as comparison for the same task. In the latter, τ and m is determined according to C-C method. Then the reconstructed data is forecasted by LSSVM. The parameters of LSSVM are optimized by GA. For the purpose of making an objective evaluation on the prediction model, RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percent Error) is introduced to evaluate the prediction model.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(y_i - \hat{y}_i \right)^2} \tag{13}$$

$$MAPE = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

(14)

where, y_i and $\hat{y}_i$ are the actual value and the forecasting value respectively.

The forecasting results from the proposed method and C-C-LSSVM are illustrated in Fig. 1, as well as the actual load curve. Overall, the forecast curve of the proposed method could fit the actual load better. Fig. 2 shows the relative error of different models. The maximum absolute relative error from proposed method is 4.61%, while that of C-C-LSSVM is as high as 4.93%. The minimum absolute relative error of C-C-LSSVM and the proposed method is 0.92% and 1.12% respectively. What's more, in C-C-LSSVM, only 58.3% of the error is below 3%. On the contrast, most error of the proposed method are satisfying, about 70.8%.

The evaluation of different methods is listed in Table 1. For the forecasting data, the MAPE of proposed method is 2.40%, which is greater in C-C-LSSVM, 2.77%. Besides, the RMSE of the proposed method and C-C-LSSVM is 5.0786 MW and 5.9991 MW, respectively.
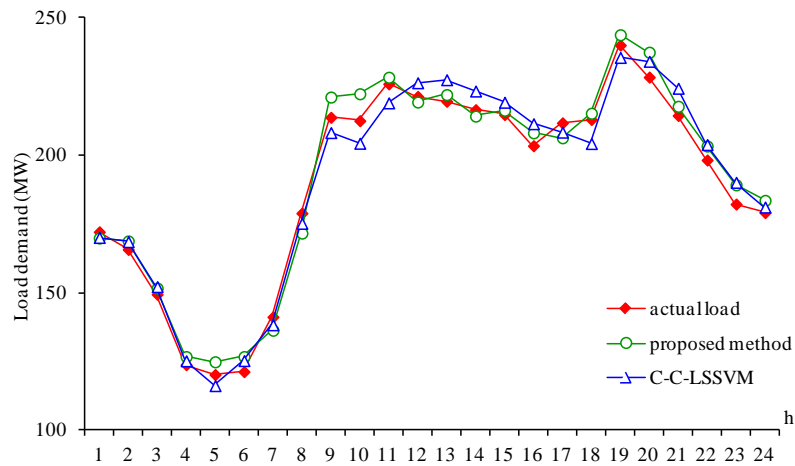

Fig. 1 Comparison between different forecasting load and actual load
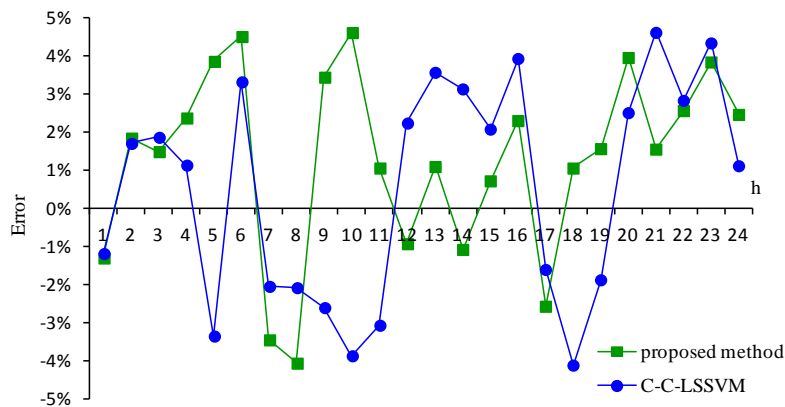

Fig. 2 Relative errors with different models

Table 1 Three Scheme comparing

Algorithm	Training		Forecasting	
	RMSE(MW)	MAPE(%)	RMSE(MW)	MAPE(%)
Proposed method	4.2519	2.47	5.0786	2.40
C-C-LSSVM	4.8287	2.95	5.9991	2.77

According to the aforementioned analysis, we can conclude that the proposed method with joint parameters optimization performs much better in the same task, and the prediction accuracy has been improved significantly.

Conclusion

The precision of LSSVM model is limited by its parameter, thus the choice of parameters has be a restricting factor to its application. This paper developed a new load forecasting scheme based on phase space reconstruction and LSSVM. The main contribution of this paper is using genetic algorithm to optimize both the parameters of phase space reconstruction and LSSVM, regarding of the overall impact to the forecasting performance. Joint optimization using GA can avoid boring manual search work. Experiment is demonstrated to show the effectiveness of this model. It can be seen from the results that the new model performance better than C-C-LSSVM whose parameters are estimated separately. In our future work, more other optimal algorithms like PSO will be tried for the joint parameter estimation. And this joint optimization can be considered in other hybrid forecasting model as well.

Acknowledgements

References

[1] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, J. Sci. Commun. 163 (2000) 51-59.

[1] G.A. Carolina, and M. Carlos, Electric power demand forecasting using interval time series: a comparison between VAR and iMLP, Energy Policy, 38 (2010) 715-725.

[2] A.Y. Alanis, L.J. Ricalde, C. Simetti, and F. Odone, Neural model with particle swarm optimization Kalman learning for forecasting in smart grids, Mathematical problems in engineering, 2013 (2013) 1-9.

[3] P. Duan, K.G. Xie, and T.T. Guo, Short-term load forecasting for electric power systems using the PSO-SVR and FCM clustering techniques, Energies, 4 (1) (2011) 173-184.

[4] S.C. Huang, and C.H. Hsieh, Wavelet-based relevance vector regression model coupled with phase space reconstruction for exchange rate forecasting, International journal of innovative computing information and control, 8 (3A) (2012) 1917-1930.

[5] D.H. Xia, S.Z. Song, and J.H. Wang, Determination of corrosion types from electrochemical noise by phase space reconstruction theory, Electrochemistry Communications, 15 (2012) 88-92.

[6] K. Unnikrishna, Prediction of magnetic substorms using a state space model, Journal of Atmospheric and Solar-Terrestrial Physics, 75-76 (2012) 22-30.

[7] M.B. Kennel, R.Brown, and H.D.I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Physical Review A, 45(6) (1992) 3403-3411.

[8] S. Boccaletti, D.L. Valladares, L.M. Pecora, H.P. Geert, and T. Carroll, "Reconstructing embedding spaces of coupled dynamical systems from multivariate data," Physical Review E, 65(3) (2002) 1-4.

[9] Y. Liu, Q.J. Shi, and J.G. Li, Study on gas emission rate prediction based on chaos analysis, Procedia Engineering, 24 (2011) 106-110.

[10] L.Y. Sun, Multivariate local polynomial regression with application to Shenzhen component index, Discrete dynamics in nature and society, 2011 (2011) 1-11.

[11] B. Qi, C.H. Zhao, and E. Youn, "Use of weighting algorithms to improve traditional support vector machine based classifications," Optics express, 19(27) (2011) 26816-26826.

[12] Q. Wu, "The hybrid forecasting model based on chaotic mapping, genetic algorithm and support vector machine," Expert systems with applications, 37(2) (2010) 1776-1983.

[13] J.A.K. Suykens, Van Gestel T., De Brabanter J., De Moor B., Vandewalle J.. Least square support vector machines. Singapore: World Scientific, 2002.

[14] S. Li, and L.K. Dai, Classification of gasoline brand and origin by Raman spectroscopy and a novel R-weighted LSSVM algorithm, Fuel, 96 (2012) 146-152.

[15] Z.H. Guo, J. Zhao, W.Y. Zhang, and J.Z. Wang, A corrected hybrid approach for wind speed prediction in Hexi Corridor of China, Energy, 36 (2011) 1668-1679.

[16] S. Ismaila, A. Shabria, and R. Samsudin, A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting, Expert Systems with Applications, 38 (2011) 10574-10578.

[17] D.E. Goldberg. Genetic Algorithms in Search, optimization and machine Learning. Addison Wesley, Reading, MA, 1989.

[18] J.C. Hung, Applying a combined fuzzy systems and GARCH model to adaptively forecast stock market volatility, Applied Soft Computing, 11 (2011) 3938-3945.

[19] M. Piltan, H. Shiri, and S.F. Ghaderi, Energy demand forecasting in Iranian metal industry using linear and nonlinear models based on evolutionary algorithms, Energy Conversion and Management, 58 (2012) 1-9.

[20] V. Sels, and M. Vanhoucke, A hybrid genetic algorithm for the single machine maximum lateness problem with release times and family setups, Computers & Operations Research, 39 (2012) 2346-2358.