

## Electricity Consumption Prediction based on Data Mining Techniques with Particle Swarm Optimization

Zeguo Qiu

*School of Computer and Information Engineering, Harbin University of Commerce  
Harbin Univ Commerce, Sch Comp & Informat Engn, Harbin 150028, Heilongjiang,  
Peoples R China  
qiuzeguo@sina.com*

### Abstract

*Accurate forecasting of electric load has always been the most important issues in the electricity industry, particularly for developing countries. Due to the various influences, electric load forecasting reveals highly nonlinear characteristics. The SVR model with Particle Swarm Optimization and Cross Validation is proposed according to the characteristics of the nonlinear electricity consumption data which are new Data Mining Techniques (DMT). In this model, PSO-CV method is used to the parameter determination. Then PSO-CV-SVR model is applied to the electricity consumption prediction of Jiangsu province. The result shows better than the ANNs method and improves the accuracy of the prediction.*

**Keywords:** *Prediction; Support vector regression (SVR); Particle Swarm Optimization (PSO); Cross validation*

### 1. Introduction

Accurate electric load forecasting can provide those export oriented economies advantages through saving and efficiently distributing limited energy resources. For inaccurate electric load forecasting, it may increase operating costs. For example, over estimation of future electric load results in unnecessary spinning reserve, wastes limited energy resources, even leads to distribution inefficiency, and, furthermore, is not accepted by international energy networks owing to excess supply. In contrast, under estimation of load causes failure in providing sufficient reserve and implies high costs in the peaking unit, which discourage any economic and industrial developments. Thus, the accuracy of future electric demand forecasting have received growing attention, particularly in the areas of electricity load planning, energy expenditure/cost economy and secure operation fields, in regional and national systems.

With the higher demand for the quality of water and the acceleration of the industrialization, the need and consumption of electricity are becoming wider. In order to meet the massive requirement of industry, and to make the economy develop smoothly, it is very important to use and program electricity power. As a result, we need to predict the electric quantity, and then figure out the appointment of production plan, thus bring out the economic and social benefits.

Support Vector Machine, rendered by Vapnik in 1995[1], can solve the phenomena of 'excess learning' with the principle of minimize the structure risk. It has great generalization capability. Applying the SVM in regression analysis, we get Support vector regression. Since the day of its birth, Experts have done lots of work to apply it into many fields and modify the

model. For instance, in paper [2], the author use immune algorithm to optimize the parameters of SVM model to predict the use of electricity in Taiwan. In paper [6] the author blend the rough set theory and the reduction of property into LS-SVM, hence improving the accuracy of prediction; Paper [3] use Ant Colony Optimization to optimize trained data and speed up SVM. While paper [7] gets satisfactory results by modeling porous NiTi alloy with SVM. Also, the paper [8] gets great results by applying the LS-SVM into prediction of financial time series.

The PSO is a non-gradient based optimization and search algorithm belonging to probabilistic search algorithms, which have attracted much attention from the research community. These algorithms generally mimic some natural phenomena, for example evolutionary algorithms (EA) and simulated annealing. EA (*i.e.*, genetic algorithms (GA), genetic programming (GP), evolutionary programming (EP), evolutionary strategies (ES), and differential evolution (DE)) model the evolution of a species, based on Darwin's principle of survival of the fittest and competition to produce better adapted generations in their problem solution space while simulated annealing is based on the statistical mechanics and models of atoms during an annealing process. The PSO in contrast to the EA, relies on cooperation rather than competition. Moreover good solutions in the problem set are shared with their less-fit brethren so that the entire population improves and poorly performing members are not killed off as in GA. Although these probabilistic search algorithms generally require many more function evaluations to find an optimal solution, as compared to gradient-based algorithms (*i.e.*, simple gradient descent, conjugate gradient descent, *etc.*), they are promising approaches due to their effectiveness in searching very large spaces and the ability to perform global search for best forecasting model. Moreover, they are generally easy to program, can efficiently make use of large numbers of processors, do not require continuity in the problem definition, and generally are better suited for finding a global, or near global, solution. In particular these algorithms are ideally suited for solving discrete and/or combinatorial type optimization problems. A survey of evolutionary algorithms in evolving artificial neural networks can be found in Ref.

In this paper, we use the consumption of electricity and macro-economy influencing data from 2004 to 2009 in Jiangsu Province. By implementing the Particle Swarm algorithm and cross validation to optimize the parameters of SVM, we construct PSO-CV-SVR. The results show that this model, with higher accuracy of prediction, is superior to BP-neural network both in fitting and error.

## 2. Principle of SVR

Assume that

$$f(x) = w\phi(x) + b \quad (1)$$

$\phi(x)$  is non-linear transformation, which convert the dataset  $x$  into high dimension characteristics space  $F$ ,  $w$  is called weight vector,  $b$  is classification threshold. We can see  $f(x)$  is a linear function of  $\phi(x)$ . Under the principle of minimum structure risk, the desired  $f(x)$  should satisfy :

$$R(f) = \frac{1}{2} \|w\|^2 + \frac{C}{N} \sum_{i=1}^n L(y_i, f(x_i)) \quad (2)$$

In which,  $C$  is a positive constant. It is a compromise factor in smooth and experience error, which is also called penalty factor.  $L(y_i, f(x_i))$  is loss function. Usually, we let the loss function be the  $\varepsilon$  non-sensitive loss function, which means for  $i = 1, 2, \dots, n$ ,

$$L(y_i, f(x_i)) = L_\varepsilon(y_i, f(x_i)) = \begin{cases} 0, & |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{ot her s} \end{cases} \quad (3)$$

Thus, the above regression solving problem becomes the optimization problem as follows:

$$\min R(w, \xi_i, \xi_i^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (4)$$

$$s.t. \begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i \\ f(x_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{cases}$$

In order to solve (4), we define Lagrange function :

$$L(w, b, \xi_i, \xi_i^*, \alpha_i, \beta_i, \mu_i, \nu_i) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

$$+ \sum_{i=1}^n \alpha_i (y_i - w\phi(x_i) - b - \varepsilon - \xi_i) + \sum_{i=1}^n \beta_i (w\phi(x_i) + b - y_i - \varepsilon - \xi_i^*) \quad (5)$$

$$- \sum_{i=1}^n \mu_i \xi_i - \sum_{i=1}^n \nu_i \xi_i^*$$

For partial derivative of  $L(w, b, \xi_i, \xi_i^*, \alpha_i, \beta_i, \mu_i, \nu_i)$  variable  $w, b, \xi_i, \xi_i^*$ , by letting the result be zero, we get:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n \alpha_i \phi(x_i) + \sum_{i=1}^n \beta_i \phi(x_i) = w - \sum_{i=1}^n (\alpha_i - \beta_i) \phi(x_i) = 0 \quad (6)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \beta_i - \sum_{i=1}^n \alpha_i = 0 \quad (7)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 \quad (8)$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \beta_i - \nu_i = 0 \quad (9)$$

Through taking (6) ~ (9) to (5), while at the same time assuming that kernel function  $K(x_i, x_j) = \phi(x_i) \bullet \phi(x_j)$ , and changing the optimization problem into its own antithesis problem, we have:

$$\max \sum_{i=1}^n y_i (\alpha_i - \beta_i) - \varepsilon \sum_{i=1}^n (\alpha_i + \beta_i) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \beta_i) (\alpha_j - \beta_j) K(x_i, x_j)$$

$$s.t. \begin{cases} \sum_{i=1}^n \beta_i - \sum_{i=1}^n \alpha_i = 0 \\ \alpha_i, \beta_i \in [0, C], i = 1, 2, \dots, n \end{cases} \quad (10)$$

To solve quadratic programming (10),

$$f(x) = \sum_{i=1}^n (\alpha_i - \beta_i) \phi(x_i) \bullet \phi(x) + b = \sum_{i=1}^n (\alpha_i - \beta_i) K(x_i, x) + b \quad (11)$$

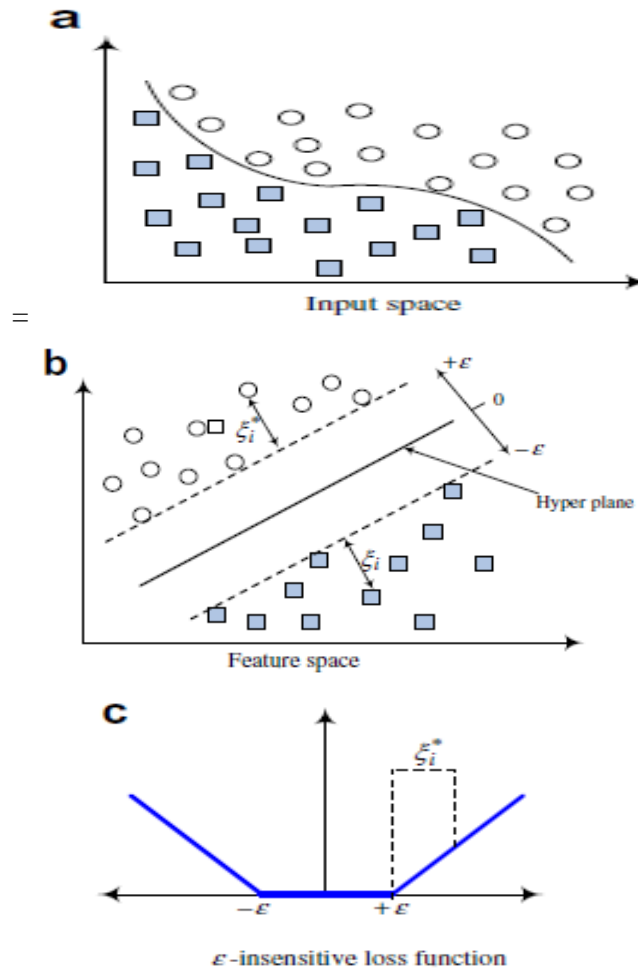
In which

$$w = \sum_{i=1}^n (\alpha_i - \beta_i) \phi(x_i) \quad (12)$$

Due to kernel function [1] of satisfying Mercer, this means corresponding to a group of dot product in high dimension space. What we have to know is the specific function that meets this condition.

Then we can get the regression function  $f(x)$  even though we have no idea of specific formulation of  $\phi(x)$ . Here we let the kernel function be radial basis function,

$$K(x, x_i) = \exp(-|x - x_i|^2 / \sigma^2) \quad (13)$$



**Figure 2.1. Transformation Process Illustration of a SVR Model**

An nonlinear mapping function  $u(x)$  defined to convert a nonlinear problem in the original (lower dimensional) data input space (a) to linear problem in a (higher dimensional) feature

space (b). The points lying on or outside the  $\epsilon$ -tube of the decision function are support vectors (b). On the right, the  $\epsilon$ -insensitive loss function is shown in which the slope is determined by  $C$  (c).

### 3. Finding of Optimization by SVR with Particle Swarm

Particle swarm optimization is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling. PSO shares many similarities with evolutionary computation techniques such as the GA. The system is initialized with a population of random solutions and searches for optima by updating generations. However, unlike the GA, the PSO has no evolution operators such as crossover and mutation. In the PSO, the potential solutions, called particles, fly through the problem space by following the current optimum particles. Each particle keeps track of its coordinates in the problem space, which are associated with the best solution (fitness) it has achieved so far (The fitness value is also stored). This value is called  $pbest$ . Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the neighbors of the particle. This location is called  $lbest$ . When a particle takes all the population as its topological neighbors, the best value is a global best and is called  $gbest$ .

The particle swarm optimization concept consists of, at each time step, changing the velocity of (accelerating) each particle toward its  $pbest$  and  $lbest$  locations (local version of PSO). Acceleration is weighted by a random term, with separate random numbers being generated for acceleration toward  $pbest$  and  $lbest$  locations. In the past several years, PSO has been successfully applied in many research and application areas. It is demonstrated that PSO gets better results in a faster, cheaper way when compared with other methods. Another reason that PSO is attractive is that there are few parameters to adjust. One version, with slight variations, works well in a wide variety of applications. Particle swarm optimization has been considered for approaches that can be used across a wide range of applications, as well as for specific applications focused on a specific requirement.

By the process of SVR algorithm, we know that if  $\epsilon$  in non-penalty function, penalty factor  $C$  and  $\sigma^2$  in radial basis function are different, then can have different model of SVR. As a result, in this paper we focus on the choosing of  $\epsilon$ , and finding the approximate optimal of  $(C, \sigma^2)$  using SVR with particle swarm algorithm, thus construct the PSO-CV-SVR model for regression forecasting.

PSO parameter selection schemes introduced in literature can be classified in three categories. In the first category, all parameters of PSO are selected empirically. Although this approach may lead to the suboptimal results for a single problem, it would not be helpful as a general approach to solve optimization problems. Algorithms in the second category try to choose time variant parameter values, e.g. adjust particle to do more exploration rather than exploitation at the beginning. Though it helps the swarm to have different behavior during its lifetime, it could vary in different applications. Finally, algorithms in the third category adaptively change PSO parameters over time by looking on the swarm state.

The crucial idea of Partial Swarm optimization to get best parameters is considering  $(C, \sigma^2)$  being the location of particles while at the same time, installing a reasonable target function. When each partial is searching location, its aim is to minimize or maximize target function, thus can locate its optimal location in history, and change in its second foundation.

Letting the target function be the means of square error function,

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (14)$$

In which  $y_i$  is the real value,  $\hat{y}_i$  is the predicted value.

We note that  $(C, \sigma^2)$  be  $x = (x_1, x_2)$ , which consists of partials of group. So the location of  $i$  th partial can be expressed as  $x_i = (x_{i1}, x_{i2})$ . Let the velocity of  $i$  th partial be  $v_i = (v_{i1}, v_{i2})$ , its historical optimal spot is  $p_i = (p_{i1}, p_{i2})$  and its overall optimal spot is  $p_g = (p_{g1}, p_{g2})$ . So the partial's location and velocity change based on the function as follows:

$$v_{ij}^{(r+1)} = wv_{ij}^{(r)} + c_1\delta_1(p_{ij}^{(r)} - x_{ij}^{(r)}) + c_2\delta_2(p_{gj}^{(r)} - x_{ij}^{(r)})$$

$$x_{ij}^{(r+1)} = x_{ij}^{(r)} + v_{ij}^{(r+1)}, \quad j = 1, 2$$

In which  $c_1$  and  $c_2$  are called learning factors, which usually are 2.  $\delta_1$  and  $\delta_2$  are pseudo-random numbers in  $[0,1]$ .  $w$  is inertia weight, whose values will influence the explore and development ability of algorithm. We let it to be time-changing weight.

Suppose

$$w \in [w_{\min}, w_{\max}],$$

$$cw_i = w_{\max} - \frac{w_{\max} - w_{\min}}{Iter\_max} * i,$$

In which  $Iter\_max$  stands for the maximum interaction times. Let  $[w_{\min}, w_{\max}] = [0.1, 0.9]$

Then we use the Cross Validation to optimize PSO-SVR model, so we can find out the more reasonable parameter set  $(C, \sigma^2)$ , thus can get smaller estimated error. The common K-CV method is to divide the sample set into  $K$  groups. The  $K-1$  groups are trained sets and the other group is the test group. Then repeat this  $K$  times. After all the  $K$  groups becoming the test set, we have  $K$  models. Finally, we use the average of the mean square errors produced by the  $K$  models.

When the improved SVR is applied to energy predicting, the pretreatment of data must first implement. Then the detailed procedures are as follows.

Step 1: Pretreatment of sample data set. Since the consumption data are in different dimension and value, normalization is implemented so as to guarantee the accuracy and the stability.

Step 2: Determine the training data set and the inputs of the testing data set.

Step 3: Set PSO parameters, including population size, the maximum number of iteration, the initial values of inertia weights, learning factors, the initial values and bounds of velocities.

Step 4: Update the position and velocity of each particle according to (1) and (2)

Step 5: Compute the fitness of each particle. Update the best individual and global positions by far.

Step 6: If the maximum number of iteration is reached, go to the next step with the obtained values of SVM parameters; otherwise, return to Step 4.

Step 7: Training the SVM with the sample data.

Step 8: If the stop criterion is satisfied, that is, the RMSE is less than a predefined value, the training process end, go to the next step with the obtained model parameters; otherwise return to Step 7.

Step 9: Predict the values of outputs using the inputs of testing and the model parameters obtained in the process of training.

Step 10: End. From the above procedures, the flowchart of improved SVM can be illustrated as follows.

The construction of the PSO-PLS-SVR model is described in detail, below:

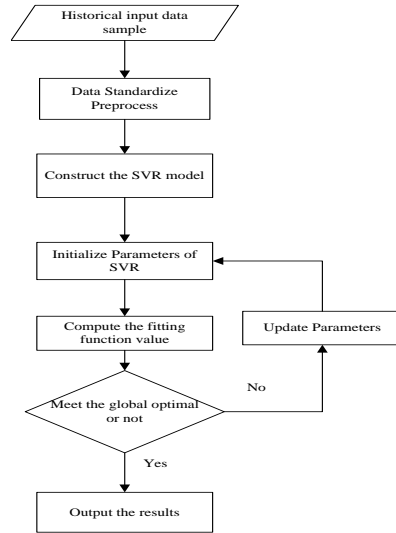


Figure 3.1. The Flow Chart of PSO-CV-SVR Model

## 4. Modeling and Prediction

### 4.1. Data Choosing and Pre-disposing

In this paper, we choose the influencing factors data and the overall electricity consumption of Jiangsu Province from January 2004 to October 2009. And we consider the data of January 2004 to July 2009 to be the trained sets, thus construct PSO-CV-SVR model. We think the data from August to October 2009 are the test sets.

Meanwhile, in order to eliminate of dimension influence, we apply the standard 0-1, that is

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

So the new data sets are all in [0,1], and the data sets also eliminate the diversity units, which interfere the results.

As the error  $\varepsilon$  of non-sensitive loss function is too small, even though it can enhance the accuracy of trained model, it reduces generalization ability. On the contrary, if  $\varepsilon$  is too bigger, the constructed model hardly can depict the change of electricity quantity. After repeated experiment, we define  $\varepsilon$  to 0.01. It can both guarantee the fitting accuracy and the generalization ability of model.

#### 4.2. Result of Regression Forecasting of the Model

Due to electricity consumption trends fluctuate by influencing factors. Thus we choose the influencing factors include Average Month Temperature (AMT), Social Retail Sales of Consumer Goods (SRSCC), Industry Increasing Value(IIV), Consumer Price Index (CPI) and Gross Value of Export-Import(GVEI). Hence, we select five influencing factors as input variable. Considering there has multi-collinearity between the influencing factors. This paper use LS to extract the main component variable. Finally, we can get the main influencing include AMT, IIV and GVEI. Then we can get the standardized data for the experimental analysis as shown in Table 4.1.

**Table 4.1. The Standardized Data of Electricity Consumption**

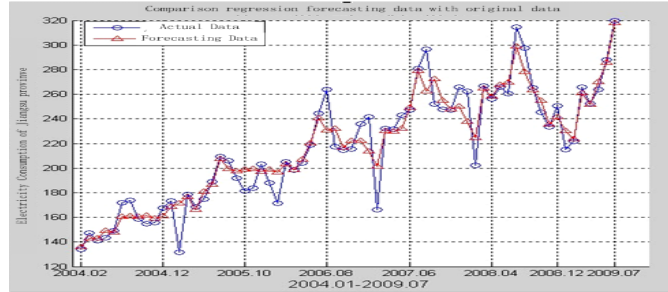
Month	Actual Data	AMT	IIV	GVEI
2007.11	-0.5499	-0.6467	-1.1757	-0.4207
2007.12	0.0625	-1.1832	-1.2434	0.1248
2008.01	-0.0437	-1.5365	-1.5636	-0.2124
2008.02	-2.0560	-0.9902	1.331	-1.5246
2008.03	0.0895	-0.6829	0.5339	0.1136
2008.04	-0.2572	-0.1138	0.2403	-0.2385
2008.05	0.0595	0.6943	1.1632	0.2638
2008.06	-0.1167	0.9105	0.7437	-0.2017
2008.07	1.6960	1.2406	-0.1792	1.2472
2008.08	1.1227	1.3658	0.5759	1.0246
2008.09	0.0481	0.7853	-0.389	0.1357
2008.10	-0.6237	0.1593	-1.1861	-0.5021
2008.11	-1.0194	-0.6487	-1.2701	-1.3251
2008.12	-0.4391	-1.1837	-1.1738	-0.5072
2009.01	-1.6331	-1.4337	-1.4782	-1.5941
2009.02	-1.4136	-0.9801	1.2635	-1.7118
2009.03	0.0658	-0.5924	0.4737	0.2031
2009.04	-0.3840	-0.1036	0.3104	-0.4170
2009.05	0.0017	0.7132	1.0946	0.2205
2009.06	0.7970	0.8903	0.8213	0.5281
2009.07	1.8568	1.1205	-0.2247	1.3639
2009.08	1.7928	1.2736	0.6168	1.5812
2009.09	0.6358	0.6902	-0.4125	0.7264
2009.10	0.3084	0.1473	-1.2536	0.2631

As the error  $\varepsilon$  of non-sensitive loss function is too small, even though it can enhance the accuracy of trained model, and it can reduce generalization ability. On the contrary, if  $\varepsilon$  is too bigger, the constructed model hardly can depict the change of electricity consumption. After repeated experiment, we define  $\varepsilon$  to 0.01. It can both guarantee the fitting accuracy and the generalization ability of model.



### 4.3. Experiment Study

In this sub- section, we construct and analyze the experiment of the proposed model in this paper using matlab tools, the predictive result of regression forecasting using PSO-CV-SVR forecasting model is showed in Figure 4.1.



**Figure 4.1. The Comparison of Forecasting Result**

Besides, we give out the regression fitting results and Relative Errors (RE) from November 2007 to July 2009, the prediction results and relative errors from August to October 2009. It is showed in the Table 4.2.2009.11

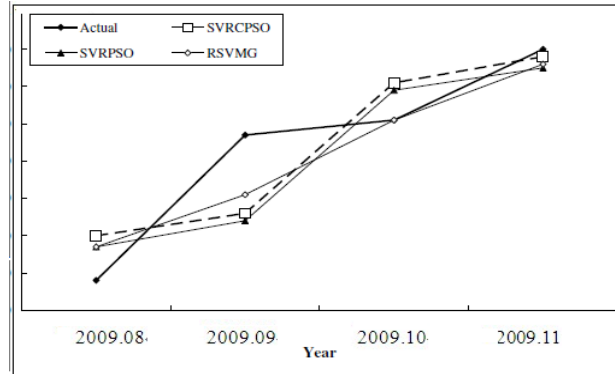
**Table 4.2. The Fitting Regression and Relative Errors**

Month	Actual Value	The fitted values Value	Relative Error
2007.11	247.42	247.6248	0.08%
2007.12	265.77	250.3068	-5.82%
2008.01	262.59	237.8922	-9.41%
2008.02	202.29	224.3531	7.91%
2008.03	266.58	264.6364	-0.73%
2008.04	256.19	258.1686	0.77%
2008.05	265.68	267.5641	0.71%
2008.06	260.4	269.3166	3.42%
2008.07	314.72	299.0822	-4.97%
2008.08	297.54	277.8595	-6.61%
2008.09	265.34	263.3971	-0.73%
2008.10	245.21	254.8402	3.93%
2008.11	233.35	235.2366	0.81%
2008.12	250.74	241.7309	-3.59%
2009.01	214.96	230.0365	7.01%
2009.02	221.54	223.4283	0.85%
2009.03	265.87	260.982	-1.84%
2009.04	252.39	251.4928	-0.36%
2009.05	263.95	270.2079	2.37%
2009.06	287.78	285.8141	-0.68%
2009.07	319.54	317.6724	-0.58%

Where the relative error is  $RE = (\hat{y}_i - y_i)/y_i$ ,  $y_i$  is the real value,  $\hat{y}_i$  is the predicted value.

From the result of Table 4.2, PSO-CV-SVR leads to a satisfactory result of electricity consumption from August to October 2009, and the relative errors confines in 10%. Moreover, the relative errors of last two months are confines to 5%.

In order to assess the rationality of the model is proposed in this paper. We compare the PSO-CV-SVR model with the PSO-SVR model and LS-SVR, the analysis result is showed as Figure 4.2.



**Figure 4.2. Relative Error Comparison with other Forecasting Method ( $\varepsilon = 0.01$ ,  $K = 3$ )**

	PSO-CV -SVR	PSO-SVR	LS-SVR
<i>MSE</i>	0.003	0.087	0.012
<i>R</i> <sup>2</sup>	94.3051%	85.3185%	89.5624%
<i>(C, σ</i> <sup>2</sup> <i>)</i>	(47.0656, 7.5959)	----	----

Where

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

$$R^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2},$$

$y_i$  is the actual value,  $\hat{y}_i$  is the predictive value,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ ,  $\bar{\hat{y}} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i$ .

*MSE* measures the deviation of predictive value from the actual value. The *MSE* is smaller shows that deviation degree is smaller, and the predictive precision of the forecasting model is more accurate.

*R*<sup>2</sup> measures the fitting degree of the forecasting model, the value of *R*<sup>2</sup> is the bigger ( $0 \leq R^2 \leq 1$ ), the fitting degree of the model is higher. Meanwhile, it reflects the extent that the dependent variables can be explained by the independent variables.

Besides, from Figure 4.2, we can see that the PSO-CV-SVR model is superior to PSO-SVR Neural Network and RSVMG both from *MSE* and fitness.

## 5. Conclusion

The research of this paper is based on SVR to construct model and predict. We solve the problems of picking parameters. We use the partial swarm optimization to find out approximate optimal value and at the same time, use the cross validation to lower predicted errors and construct PSO-CV-SVR model. By doing this, on one hand, SVR can solve the non-linear, big volatility problem. On the other hand, the parameters choosing problems can also be solved by partial swarm optimization. From the final result, we can draw a conclusion confidently that the model we build, with higher accuracy, is superior to BP-neural network

## Acknowledgements

This work was partially supported by bellows: First, “low voltage power line OFDM system”(item number: F200914), Natural Science Foundation of Heilongjiang Province. Second, “Research on innovation ability cultivation of postgraduate based on construction of scientific and innovation team in universities” (item number: C201200059), Research Projects in Education Teaching of Harbin University of Science and Technology. Third, “New type of device for fault current detection”, University Students' Innovative Undertaking Training Program of Harbin University of Science and Technology in 2013. Fourth, “Research on Torque rheometer remote monitoring system” (item number: 2009RFQXG072), Youth fund of Harbin University of Science and Technology.

## References

- [1] S. Oh and J. Kwak, “Mutual Authentication and Key establishment mechanism using DCU certificate in Smart Grid”, Applied Mathematics & Information Sciences, Mag., vol. 1S, (2012), pp. 257-264.
- [2] V. N. Vapnik, “The Nature of Statistic Learning Theory”, New York: Springer, (1995).
- [3] W.-C. Hong, “Electric load forecasting by support vector model”, Applied Mathematical Modelling, vol. 33, (2009), pp. 2444-2454.
- [4] D. Niu and Y. Wang, “Power load forecasting using support vector machine and ant colony optimization”, Expert Systems with Applications, vol. 37, (2010), pp. 2531-2539.
- [5] Y. Shi and R. Eberhart, “A Modified Particle Swarm Optimizer”, International Conference on Evolutionary Computation Pro-ceedings, (1998), pp. 69-73.
- [6] K. Kavaklioglu, “Modeling and prediction of Turkey’s electricity consumption using Support Vector Regression”, Applied Energy, vol. 88, (2011), pp. 368-375.
- [7] A. Goia, C. May and G. Fusai, “Functional clustering and linear regression for peak load forecasting”, International Journal of forecasting, Mag. vol. 26, (2010), pp. 700-711.
- [8] D. Akay and M. Atak, “Grey prediction with rolling mechanism for electricity demand forecasting of Turkey”, Energy, Mag. vol. 32, (2007), pp. 1670-1675.
- [9] N. Amjady and F. Keynia, “A New Neural Network Approach to Short Term Load Forecasting of Electrical Power Systems”, Energies, Mag., vol. 4, (2011), pp. 488-503.
- [10] M. Barker and W. Rayens, “Partial least squares for discrimination”, Journal of Chemometrics, Mag. vol. 17, (2003), pp. 166-173.
- [11] V. N. Vapnik, “The Nature of Statistic Learning Theory”, New York, Springer, (1995).
- [12] K. Kim, “Financial time series forecasting using support vector machines”, Neurocomputing, Mag. vol. 55, (2003), pp. 307-319.
- [13] J. Wang, W. Zhu, W. Zhang and D. Sun, “A trend fixed on firstly and seasonal adjustment model combined with the  $\epsilon$ -SVR for short-term forecasting of electricity demand”, Energy Policy, Mag. vol. 37, (2009), pp. 4901-4909.
- [14] P.-F. Pai, K.-P. Lin and C.-S. Lin, “Time series forecasting by a seasonal support vector regression model”, Expert Systems with Applications, Mag. vol. 37, (2010), pp. 4261-4265.
- [15] K. Kavaklioglu, “Modeling and prediction of Turkey’s electricity consumption using Support Vector Regression”, Applied Energy, Mag. vol. 88, (2011), pp. 368-375.
- [16] Q. Zhang, M. Liu, Z. H. Ling and M. Gao, “Design of Energy Consumption Prediction System of Iron and Steel Enterprises”, Metallurgical Power, vol. 2, (2006), pp. 67-70.

- [17] S. R. Valsalam, V. Muralidharan and N. Krishnan, "Implementation of energy management system for an integrated steel plant", Proceedings of Energy Management and Power Delivery, vol. 2, (1998), pp. 661-666.
- [18] F. John, "Applied Regression Analysis", Linear Models and Related Methods, Sage, (1997).
- [19] E. P. B. George and J. Gwilym, "Time series analysis, forecasting and control", Holden Day, Incorporated, (1990).
- [20] M. B. Christopher, "Neural networks for pattern recognition", Oxford University Press, (1995).
- [21] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers", Neural Process. Lett. (S1370-4621), vol. 9, no. 3, (1999), pp. 293-300.
- [22] O. Chapelle, V. Vapnik, O. Bousquet and S. Mukherjee, "Choosing multiple parameters for support machines", Mach. Learn, vol. 46, (2002), pp. 131-159.
- [23] R. Eberhart and J. Kennedy, "A New Optimizer Using Particle Swarm Theory", Sixth International Symposium on Micro Machine and Human Science, (1995), pp. 39-43.
- [24] Y. H. Shi and R. Eberhart, "A modified particle swarm optimizer", Proceedings of IEEE International Conference on Evolutionary Computation, Anchorage, (1998), pp. 69-73.

## Author



**Zeguo Qiu**, he once majored in Computer Science and Technology (2004). Now he is a doctor candidate majoring in Management and Decision with Information Systems (2011) and a teacher in Computer and Information Engineering College of Harbin University of Commerce. His current research interests include computer information system, management decision support, data mining and BPR.