The Optimized Comparison of the Gray Model Improved by Posterior-Error-Test and SVM Modified by Markov Residual Error in the Long-medium Power Load Forecast

Li wei
Department of Economics and Management, North China Electric Power University, Baoding, Hebei, 071003, China
hd11111@126.com

Zhang zhen-gang
Department of Economics and Management, North China Electric Power University, Baoding, Hebei, 071003, China
zhangzhengangwf@163.com

Abstract — **Generally, the long-medium power load forecasting sequence has small sample, stochastic growth and nonlinear wave characteristics. Gray and SVM model could reflect the relationship between growing characters- -tics and nonlinear characteristics to the series effectively and make fitting calculation. The paper modifies the proposed gray model through posterior-error-test and compares the predictive value of power load when the evaluation result is best with the optimal result forecast by SVM that is modified by Markov residual. Then we can find which model is the better. As result, we can see that Markov could well reflect randomness that produced by the system involve with many complex factors. A forecast model based on SVM algorithm is established, the series of historical load variables is rolling forecasted. It is proved that the presented forecast method is superior obviously to traditional methods through empirical study, and it can be used generally.**

Index Terms —**Posterior-error-test, GM(1,1) model, Markov, SVM residual error, power load forecast**

I. INTRODUCTION

Long-medium power load forecast is the necessary condition and foundation for the program, design, researching, production and operation of power system, the accurate result of the load forecast has an direct influence on the capital management of power construction. Especially, the power supply and consumption which are lead by various undetermined factors is surge. The power load will have excellent effects.

The gray SVM method mentioned in [4][5] gained favorable results. Using the combined weight reveal the future development trend of system. So the residual error modified method was used in power load forecasting. The credibility and precision can be further improved in practice, but lack more modification to errors. The posterior-error-test and weighted Markov residual error method [6] is presented to modify forecast results, and it has a good effect. According to an appropriate normalization method, we consider the alterative value of historical load as input variables for GM(1.1) or SVM method, and get GM(1.1) or SVM load forecast value by the way of intelligent optimization forecast. At last, we obtain the final forecast results through the improved Markov process to modify the forecast error, the final results are tested by empirically and an ideal result.

II. GM(1,1) MODEL

GM (1, 1) model is one of the common gray models, it consists by the single variable differential equation, and it is the effective model for power load forecasting.

Supposing there is a data list named $x^{(0)}$

$$x^{(0)} = \left[x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\right] \tag{1}$$

Using 1-AGO to create the progression list

$$x^{(1)} = \left[x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\right]$$

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)\left(k = 1, 2, \cdots, n\right) \tag{2}$$

$x^{(1)}$ According to the model as follows

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \tag{3}$$

Using the least square method to make the parameter $\hat{a}, \hat{u}$

$$\hat{A} = (B^T B)^{-1} B^T Yn = \begin{pmatrix} \hat{a} \\ \hat{u} \end{pmatrix}$$

doi:10.4304/jcp.5.5.703-708

$$Y_n = \begin{pmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{pmatrix} B = \begin{pmatrix} -\frac{1}{2}\left[x^{(1)}(1)+x^{(1)}(2)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(2)+x^{(1)}(3)\right] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left[x^{(1)}(n-1)+x^{(1)}(n)\right] & 1 \end{pmatrix} \quad (4)$$

Sending $\hat{a},\hat{u}$ to the differential equations

$$x^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}}, (k = 0,1.2...) \quad (5)$$

Progression decrease the result to $x^{(0)}$

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$$
$$= (1-e^{\hat{a}})(x^{(0)}(1) - \frac{\hat{u}}{\hat{a}})e^{-\hat{a}k}, (k = 0,1,2...) \quad (6)$$

III. SVM ALGORITHM

SVM(Support vector machines) method is the effective practice to statistical learning theory, it is based on VC theory and Structural Risk Minimization, the optimum fitting between model complexity and learning ability based on finite sample information was established to reach the best generalization, it significantly better than the neural network method based on empirical risk minimization,it also can better solve the practical problems in the small sample, nonlinear and high dimensional. It has simple result, global optimization and better generalization ability in characteristics:

(1)It realizes the SVM principle, also minimizes the generalization error upper boundary, so it has the better upper boundary ability.

(2) Compared with neural network method, SVM has the less free parameters. There are three free parameters in SVM algorithm, but it should be selected by subjective.

(3)Neural network cannot always meet the global optimal solution, and it can easily fall into local optimal solution. In SVM algorithm, training SVM equivalent to solute a two convex planning problem with nonlinear constraint,so the solution is unique, global and optimal[1-3].

The disadvantage of SVM is that it can't determine which knowledge is redundancy. The core theory of SVM is making the input variables map to the high dimensional feature space H by nonlinear mapping $\varphi(\bullet)$, the optimum decision function is established by structure risk minimization principle, then the kernel function in original space is used to replace the dot product operation in high dimensional feature space.The optimum regression function in high dimension space H for SVM is as follows:

$$f(x) = w \bullet \varphi(x) + b \quad (7)$$

w——vector, $w \in R^k$

b——intercept, $b \in R$

At present, SVM includes ε-SVR and v-SVR. This paper used the v-SVR to explain. The v-SVR can be transformed into an optimization problem:

$$\frac{\min}{w,b,\xi_i^-,\xi_i^+}\frac{1}{2}w^T w + C\left[v\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i^- + \xi_i^+)\right]$$

$$s.t.\begin{cases} y^i - f(x^i) \le \varepsilon + \xi_i^- \\ f(x^i) - y^i \ge \varepsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \ge 0 \quad i = 1,\cdots,l \end{cases} \quad (8)$$

C is penalty factor, v is the number of support vector machine, ε is insensitive loss function, $i = 1,\cdots,l$

The dual problem is as follows:

$$\frac{\min}{\alpha,\alpha^*}\frac{1}{2}(\alpha-\alpha^*)^T Q(\alpha-\alpha^*) + y^T(\alpha-\alpha^*)$$

$$s.t.\begin{cases} e^T(\alpha-\alpha^*) = 0 \\ e^T(\alpha+\alpha^*) \le Cv \\ o \le \alpha_i, \alpha_i^* \le C/l \quad i = 1,\cdots,l \end{cases} \quad (9)$$

$Q_{ij} = K(x_i,x_j) \equiv \varphi(x_i)^T \varphi(x_j), K(x,y)$ is kernel function, $\varphi(\bullet)$ is mapping function in high dimension space.

The common kernel functions are as follows:

(1) linear kernel function

$$k(x,y) = (x,y)$$

(2) polynomial kernel function

$$K(x,y) = (x \bullet y + 1)^d, d = 1,2,\cdots$$

(3) RBF kernel function

$$K(x,y) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right)$$

(4) Laplace Function Kernel

$$K(x,y) = \prod_{i=1}^{n} e^{-u|x_i - y_i|}$$

(5) Radial Basis Function Kernel

$$K(x,y) = e^{-g\cdot\sum_{i=1}^{n}(x_i-y_i)^2}$$

(6) Cauchy Function Kernel

$$K(x,y)=\prod_{i=1}^{n}\frac{1}{1+u\left|x_i-y_i\right|^2}$$

$$u\in R^+$$

The approximation regression function is

$$f(x)=\sum_{i=1}^{l}\left(-\alpha_i+\alpha_i^*\right)K(x_i,x)+b \qquad (10)$$

The SVM structure is showed in Fig.1, $\alpha_i-\alpha_i^*$ is neural weight, $x_1,x_2,\cdots,x_m$ is input vector, y is output vector.

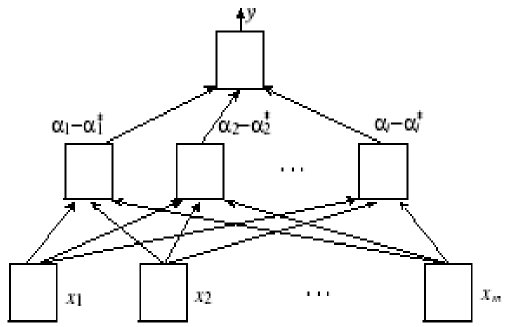

Figure1. the structure of SVM

IV. The Improved Markov Model

As the state probability that stochastic process on t_0 is known, the state probability more than t_0 only related to t_0, but it independent the state probability before t_0, we called this process is Markov process.

Assumed the state space of random sequence $\{x(n), n=1,2,...\}$ is E, if any integer $n_1,n_2,...,n_m\left(0\le n_1<n_2<...<n_m\right)$ and any natural number k meet

$$p\{X(n_m+k)\}=p\{X(n_m+k)=j\,|\,X(n_m)=i_m\}$$

$i_1,i_2,..i_m, j\in E$, then random sequence $\{x(n), n=1,2,...\}$ is markov chain. It can be showed as follows:

$$x(n)=x(0)p^n,\ p=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}$$

$x(n)$ is the state probability vector in time n, $x(0)$ is the state probability vector in initial time, p is probability transfer matrix，the each row is a probability vector, it shows the probability that system state E_1 of i row transfer to other states[10-11].

Based on calculation, we can get the relative error (%) δ, according to δ, the four status E1, E2, E3, E4 are as follows:

Status	Value limit
E1	$\delta=(0,1]$
E2	$\delta=(-1,0]$
E3	$\delta=(1,6]$
E4	$\delta=(-6,\ -1]$

Suppose the total sample numbers is N, the transition probability from p to q is

$P_{pq}=\frac{N_{pq}}{N}$, N_{pq} is the transfer times from p to q

The probability status of system error can be determined by markov chain:

$\delta^*=\frac{\delta_{up}+\delta_{down}}{2}$, the power load forecasting value $F=F\bullet\left(1\pm\delta^*\right)$.

Due to the optional determination of divide upper and lower bound in an interval by the traditional Markov modification method, the problem of errors corrected without measure often occurs. So we proceed with selection method of upper and lower bound, round the error of one position before forecast position, and δ_{up} is supposed to be the absolute value of the rounded error, δ_{down} equals to 0.01, the sign before δ^* is determined by the state interval. With these works, the error modification can be more accurate, it is put into practice and turns out to be good effect.

V. Demonstration

In order to verify the forecast effect of the proposed method, the historical data of power consumption for a city is used to predict by Gary and SVM method, and then modify with posterior-error-test and Markov, the forecast results are gained, and a comparative verification was made at last. Details are showed as table 1.

TABLE I.
THE ELECTRICITY CONSUMPTION

Year	Electricity consumption Billon kwh	Year	Electricity consumption Billon kwh
1995	40	2002	56.3
1996	40.8	2003	59.9
1997	49	2004	
1998	52	2005	
1999	56.9	2006	
2000	58.8	2007	
2001	56	2008	

A. Gray forecasting

The data from 1995 to 2003 are used as training samples, the data are normalized firstly, 5 former years data of power consumption are set to be input samples. The load data of 6 year which as output samples is used to rolling training by GM(1.1) method.

$$\hat{a}=-0.039131, \hat{u}=44.19$$

$$X(k+1)=(1-e^{\hat{a}})(x^{0}(1)-\frac{\hat{u}}{\hat{a}})e^{-\hat{a}k}$$

Obtained:
x(9+1)=63.82
x(10+1)=66.36
x(11+1)=69.01
x(12+1)=71.76
x(13+1)=74.63

Do Posterior-error-test to the above model, the process is:

mean residual:

$$\bar{\varepsilon}=\frac{1}{n}\sum_{k=1}^{n}\varepsilon(k)=\frac{1}{n}\sum_{k=1}^{n}\left[x^{(0)}(k)-\hat{x}^{(0)}(k)\right]$$

error variance:

$$S_1^2=\frac{1}{n}\sum_{k=1}^{n}\left[x^{(0)}(k)-\bar{x}\right]^2$$

Posterior-error-ratio C:

$$C=\frac{S_2}{S_1}$$

Infinitesimal error probability P:

$$P=P\{|\varepsilon(K)-\bar{\varepsilon}|$$

Then we utilize Posterior-error-ratio c, infinitesimal error probability P to evaluate the current model:

C=0.4621 evaluation result：good
p=0.8750 evaluation result：good
The predictive value of the next 5 times:
X(t+1)=63.81506
X(t+2)=66.36172
X(t+3)=69.01002
X(t+4)=71.76400
X(t+5)=74.62788
Qmin=-5.86244

TABLE II.
THE FORECAST ELECTRICITY CONSUMPTION CONTRAST

Serial number	True value	Predictive value	Fitting value	error
X(2)	40.80	46.66	-5.86	-14.37
X(3)	49.00	48.52	0.48	0.97
X(4)	52.00	50.46	1.54	2.96
X(5)	56.90	52.47	4.43	7.78
X(6)	58.80	54.57	4.23	7.20
X(7)	56.00	56.75	-0.75	-1.33
X(8)	56.30	59.01	-2.71	-4.82
X(9)	59.90	61.37	-1.47	-2.45

Continue to model residual series. The first analysis results of residual series:

Model parameters
$\hat{a}$=-0.155623 $\hat{u}$=2.083293
x(t+1)=19.249261exp(0.155623t)-13.386825

TABLEIII.
THE FORECAST ELECTRICITY CONSUMPTION CONTRAST TABLEII

Serial number	True value	Predictive value	Fitting value	error
X(2)	40.80	44.04	-3.24	-7.94
X(3)	49.00	52.31	-3.31	-6.76
X(4)	52.00	54.41	-2.41	-4.64
X(5)	56.90	55.63	1.27	2.23
X(6)	58.80	54.17	4.63	7.87
X(7)	56.00	53.13	2.87	5.12
X(8)	56.30	57.33	-1.03	-1.83
X(9)	59.90	63.79	-3.89	-6.49

The evaluation of the current model
C=0.4250 evaluation result：good
p=0.8750 evaluation result：good
The predictive value of the next 5 times:
X(t+1)=69.20921
X(t+2)=73.65131
X(t+3)=78.51421
X(t+4)=83.85570
X(t+5)=89.74280
Qmin=-8.96337
Continue to model residual series.
The 2nd analysis results of residual series:
Model parameters
$\hat{a}$=-0.117045 $\hat{u}$=1.645604
x(t+1)=23.022937exp(0.117045t)-14.059566

TABLEIV.
THE FORECAST ELECTRICITY CONSUMPTION CONTRAST TABLEIII

Serial number	True value	Predictive value	Fitting value	error
X(2)	40.80	37.94	2.86	7.02
X(3)	49.00	49.80	-0.80	-1.64
X(4)	52.00	55.61	-3.61	-6.95
X(5)	56.90	59.69	-2.79	-4.91
X(6)	58.80	57.47	1.33	2.27
X(7)	56.00	52.37	3.63	6.49
X(8)	56.30	54.34	1.96	3.49
X(9)	59.90	62.54	-2.64	-4.41

The evaluation of the current model

C=0.3745 evaluation result：good
p=1.0000 evaluation result：very good
The predictive value of the next 5 times:
X(t+1)=67.53763
X(t+2)=72.88516
X(t+3)=78.76591
X(t+4)=85.25164
X(t+5)=92.42505
Qmin=-4.34446
Continue to model residual series.
The 3rd analysis results of residual series:
Model parameters
$\hat{a}$ =0.039097 $\hat{u}$ =5.098622
x(t+1)=-126.066exp(-0.039097t)+130.41

TABLE V
THE FORECAST ELECTRICITY CONSUMPTION CONTRAST TABLEIV

Serial number	True value	Predictive value	Fitting value	error
X(2)	40.80	38.43	2.37	5.82
X(3)	49.00	47.24	1.76	3.58
X(4)	52.00	53.68	-1.68	-3.23
X(5)	56.90	61.20	-4.30	-7.55
X(6)	58.80	61.60	-2.80	-4.76
X(7)	56.00	55.01	0.99	1.77
X(8)	56.30	53.19	3.11	5.52
X(9)	59.90	59.28	0.62	1.03

The evaluation of the current model
C=0.3514 evaluation result: good
p=1.0000 evaluation result: very good
The predictive value of the next 5 times:
X(t+1)=66.72859
X(t+2)=71.94056
X(t+3)=77.69095
X(t+4)=84.05132
X(t+5)=91.10418
Qmin=-4.64647

B. SVM forecasting

Likewise，selecting the data from 1995 to 2003 are used as training samples, the data are normalized firstly, The load data of 6 year which as output samples is used to rolling training by SVM method, inner product function RBF is adopted to be kernel function, parameters are set as this: σ^2=30, C=100, ε=0.001. Through LIBSVM algorithm, we get load forecast values from 2004 to 2008. According to the former 5 years' data and their influencing factors, we select seven years' data from 2002 to 2008 in the table below. They are showed as table VI. .

TABLE VI.
ELECTRICITY CONSUMPTION PREDICTION RESULTS BY SVM

Year	Real data	SVM forecasting value	Relative error	status
2002	62.7	61.32	-0.022	E4
2003	65.914	62.59	-0.0504	E4
2004	59.13	60.13	0.01691	E1
2005	62.22	60.29	-0.031	E3
2006	74.812	72.91	-0.0254	E4
2007	86.686	82.21	-0.0516	E4
2008	88.64	89.43	0.00891	E3

Step one: Error analysis
We adopt relative error to analyze forecast errors:
$RE = \dfrac{y_i^* - y_i}{y_i}$, calculation results are showed as table VI.

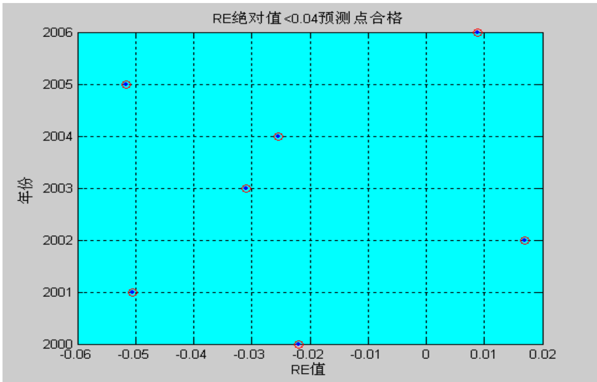

Figure2. Combination prediction relative errors

We determine qualified rate of forecasting mainly with the standard of 4% (it is a qualified forecasting position while $|RE|$ <4%), and determine forecast precision by average relative error simultaneously. The calculation results are obtained as this: average error of SVM is 0.0221; qualified rate is 71.43%, so they belong to a high level. We can see that the absolute value of RE to the forecast value are mainly between 0 and 0.04 from figure 2, this further illustrates the feasibility of the proposed method. For further optimization of forecast results, now introduce an improved Markov process to modify them.

Step two: The markov correcting

The transition probability matrix devised by relative error statement of SVM forecasting is

$$p = \begin{bmatrix} 0 & 0 & 0.17 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.17 \\ 0.17 & 0 & 0.17 & 0.32 \end{bmatrix}$$

Take the load data of 2006 for an example, according to the state probability vector of 2005 as $(1 \quad 0 \quad 0 \quad 0)^t$, using the formula we get the state probability vector of 2005, $(0 \quad 0 \quad 0.17 \quad 0)^t$, it is obviously that the largest probability of load forecast error in state of E3 is 0.17, calculating the forecast value of 2005 in formula,

$$F = F \bullet (1 + \delta^*) = 60.29 * (1 + 0.01) = 60.9 \text{ Billon kwh}$$

The monitoring load data of 2005 is 62.22 billion kWh, the value corrected by Markov is 60.9 billion kWh, and the value forecasted by SVM method is 60.29 billion kWh. It is proved that the adoption of Markov error correction method can obviously increase forecast precision, and ensure the stability of forecasting.

Select the best value of gray prediction when p gradually reaches 1.0000 through posterior-error-test:

The result of forecast load which is 71.9 billion kWh in 2005 comparing with the true value 62.22 billion kWh，we can find that the result of the modified SVM is more approached to the true value comparing with the gray prediction of Posterior-error-test

VI. CONCLUSION

The method of Posterior-error-test is utilized to optimize gray model, and the optimization of C, P are obtained. Under these conditions, we can obtain the best predictive value.

If we have the same impact factors, we can obtain the fittest values which have a min error compare with the real results. The values are got by the SVM model which modified by Markov residual error.

Analyzing the reasons of comparatively large error on gray model, we can take the following improving measures:

（1）Further optimizing original series, such as exponential weighting, moving average;

（2）Select optimization data of C, such as median value, or final value;

（3）improving model: for example, we use other methods to define but not use the median of x(0)(k) and x(0)(k+1).

Sum up, we use the method of SVM to predict the long-medium power load forecast. But there are some uncertain factors of parameters, such as C, other variables. Therefore, the problem which how to improve the intelligent algorithm must be studied.

REFERENCES

[1] Niu Dongxiao, Liu Da. “Support Vector Machine Models Optimized by Genetic Algorithm for Hourly Load Rolling Forecasting”. *Transactions of China Electro technical Society*, 2007, (6), pp.148-153.

[2] Niu Dongxiao,Gu Zhihong.etc. “Study on Forecasting Approach to Short-term Load of SVM Based on Data Mining” [J]. *Proceedings of the CSEE*, 2006, (9), pp.6-12.

[3] Yang Jinfang,Zhai Yongjie. “Time series prediction based on support vector regression”.*Proceedings of the CSEE*, 2005,(9),pp.110-114.

[4] O. Habiballah, R. Ghosh-Roy and M. R. Irving. “Markov chains for multipartitioning large power system state estimation networks”.*Electric Power Systems Research*, 1998, (3),pp. 135-140.

[5] D. I. Jones and M. H. Lorenz. “An application of a Markov chain noise model to wind generator simulation”.*Mathematics and Computers in Simulation, 1986,(8),pp.391-402.*

[6] Dong Jizheng,Wang Huan.etc. “Application of Markov chain with weights to load forecasting”[J]. *Relay*, 2006,(3),pp.33-36.

[7] Niu, Dongxiao,Li, Jinchao. “Distribution center location decision-making based on PCA and support vector machine approach,”*Conference on Natural Computation*, 2007, pp.553-557.

[8] Niu, Dongxiao,Liu, Da,Chen, Guangjuan,Feng, Yi. “Support vector machine models optimized by genetic algorithm for hourly load rolling forecasting,” *Transactions of China Electrotechnical Society*, 2007, pp.148-153.

[9] Liu,Da,Niu,Dongxiao,Xing,Mian.“Day-ahead price forecast with genetic-algorithm-optimized support vector machines based on GARCH error calibration,”*Automation of Electric Power System 2007*, pp.31-34.

[10] Liu,Da,Niu,Dongxiao,Xing,Mian. “Day-ahead price forecast with genetic-algorithm-optimized support vector machines based on GARCH error calibration”. Dianli Xitong Zidonghua/Automation of Electric Power Systems, v 31, n 11, Jun 10, 2007, p 31-34+58

[11] NIU Dong-xiao etc. “Grey load forecasting models based on relational analysis of multi-factor”, Journal of North China Electric Power University，2006.(3):91－92.

Li Wei: (1968.12.14--), Gaobeidian, Hebei province, graduate from Heibei university at Baoding Hebei province in 1991, undergraduate from north china electric power university at Baoding Hebei province in 2006,D.candidate in north china electric power university at Beijing in 2007, major field of Technical and economic management, Research Area is power load forecast .

[1]Application of Improved Ant Colony Clustering in the Power Load Forecasting, journal Beijing institute of technology, 2007.12

[2]A New Intelligent Method for Power Load Forecasting, ISDA2006,2006

[3]Improved Genetic Algorithm–GM(1,1) for Power Load Forecasting Problem, DRPT2008,2008

Current and previous research interests is power forecast load.

Zhang Zhen-gang: (1977.03.25--), Xingtai, Hebei province, graduate from north china electric power university at Baoding, Hebei province in 2002, Postgraduate in north china electric power university at Baoding Hebei province in 2008, major field of Technical and economic management, Research Area is power load forecast .