*I. J. Computer Network and Information Security*, 2012, 1, 24-30
Published Online February 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2012.01.03

# Optimization Modeling for GM(1,1) Model Based on BP Neural Network

Deqiang Zhou
*School of Information and Mathematics, Yangtze University, Jing zhou, China*
*Email: zdqmfk@yahoo.com.cn*

*Abstract*—**In grey theory, GM(1,1) model is widely discussed and studied. The purpose of GM(1,1) model is to work on system forecasting with poor, incomplete or uncertain messages. The parameters estimation is an important factor for the GM(1,1) model, thus improving estimation method to enhance the model forecasting accuracy becomes a hot topic of researchers. This study proposes an optimization method for GM(1,1) model based on BP neural network. The GM(1,1) model is mapped to a BP neural network, the corresponding relation between GM(1,1) model parameters and BP network weights is established, the GM(1,1) model parameters estimation problem is transformed into an optimization problem for the weights of neural network. The BP neural network is trained by use of BP algorithm, when the BP network convergence, optimization model parameters can be extracted, and the optimization modeling for GM(1,1) Model based on BP algorithm can be also realized. The experiment results show that the method is feasible and effective, the precision is higher than the traditional method and other optimization modeling methods.**

*Index Terms*—**Grey system; GM(1,1) model; BP neural network; Data fitting; Optimization modeling**

## 1. Introduction

Grey system theory is an interdisciplinary scientific area that was first introduced in early 1980s by Deng[1,2]. Since then, the theory has become quite popular with its ability to deal with the systems that have partially unknown parameters[3-5]. In the field of information research, deep or light colors represent information that is clear or ambiguous, respectively. Meanwhile, black indicates that the researchers have absolutely no knowledge of system structure, parameters and characteristics, while white represents that the information is completely clear. Colors between black and white indicate systems that are not clear, such as social, economic or weather systems. The fields covered by grey theory include systems analysis, data processing, modeling, prediction, decision making and control. The grey theory mainly works on systems analysis with poor, incomplete or uncertain messages. Because the grey system model needs little origin data, has simple calculate process and higher forecasting accuracy, it has been widely used in the prediction of a lot of research fields. A grey prediction model is one of the most important parts in grey system theory, and that, the GM(1,1) model is the core of grey prediction[6]. The purpose of GM(1,1) model is to work on system forecasting with poor, incomplete or uncertain messages. The GM (1,1) has more advantages with contrast to those traditional prediction ways, because it does not need to know whether the prediction variables obey normal distribution or not, does not require too much statistic sample[1]. However, many scholars find that there are many theory defects in traditional GM (1,1) model[7] , and do a lot of researches for this. These researches almost take traditional GM (1,1) modeling steps and thoughts, thus there are some following defects, first, these improved models need to face the reasonable selection of background value. On the other hand, the parameters estimation is an important factor for the GM(1,1) model, thus improving estimation method to enhance the model forecasting accuracy becomes a hot topic of researchers. In above improved models, the parameters estimation theoretical basis of transformation from discrete form to continuous form also cannot avoid the jumping errors from the difference equation to differential equation. Generally, from the evaluation criteria of the model fitting, when solving GM(1,1) model, we should directly take minimizing the error of the primitive value and the predicted value and the actual value as the criterion[9].

According to the above describing, we need find a new method to improve the precision of GM(1,1) model. From the data fitting's viewpoint, this study uses BP neural network model to improve the precision of GM(1,1) model. The grey system's primitive discrete data is fitted by a continuous model which has the same form with GM(1,1) model's time response type, the model parameters are trained and optimised by means of BP algorithm, then the direct and optimization modeling for GM(1,1) model is realized.

The remainder of this paper is organized as follows: in Section 2, the relationships between GM(1,1) model parameters and BP network weights is established, where BP network algorithm is used to optimise GM (1,1) model parameters. The experiment results and discussions are presented in Section 3. Conclusion and future work are given in the final section.

## 2. GM(1,1) model and BP neural network mapping relationships

### 2.1 GM(1,1) model

GM(1,1) type of grey model is the most widely used in the literature, pronounced as ‘‘grey model first order one variable’’. This model is a time series forecasting model. The differential equations of the GM(1,1) model have time-varying coefficients. In other words, the model is renewed as the new data become available to the prediction model. The GM(1,1) model constructing process is described below:

Consider a single input and single output system. Assume that the time sequence $X^{(0)}$ represents the outputs of the system:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)), n \geq 4 \qquad (1)$$

where $X^{(0)}$ is a non-negative sequence and $n$ is the sample size of the data. When this sequence is subjected to the accumulated generating operation (AGO), the following sequence $X^{(1)}$ is obtained.

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)) \qquad (2)$$

where $x^{(1)}(k) = \sum_{j=1}^{k} x^{(0)}(j), (k = 1,2,\cdots,n)$ .It is obvious that $X^{(1)}$ is monotonically increasing.

The generated mean sequence $Z^{(1)}$ is defined as:

$$Z^{(1)} = (z_2^{(1)}, z_3^{(1)}, \cdots, z_n^{(1)})$$

where

$z^{(1)}(k) = \lambda x^{(1)}(k) + (1-\lambda) x^{(1)}(k-1)), \quad (k = 2,3,\cdots,n)$ , generally $\lambda = 0.5$ .

The GM(1,1) model can be constructed by establishing the whitening equation for $X^{(1)}$ as:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \qquad (3)$$

Let $\hat{x}^{(1)}(1) = x^{(1)}(1) = x^{(0)}(1)$ , the solution of (3) is

$$\hat{x}^{(1)}(t) = (x^{(0)}(1) - \frac{\hat{b}}{\hat{a}}) e^{-\hat{a}(t-1)} + \frac{\hat{b}}{\hat{a}} \qquad (4)$$

In above, $[\hat{a}, \hat{b}]^T$ is a sequence of parameters that can be found as follows:

The grey difference equation of GM(1,1) is defined as follows[1,2]:

$$x^{(0)}(k)+az^{(1)}(k)=b \qquad (5)$$

The least square estimate sequence of the Eq. (5) can be found as follows:

$$[\hat{a},\hat{b}]^T=(B^TB)^{-1}B^TY \qquad (6)$$

where

$$Y=\begin{bmatrix}x_2^{(0)}\\x_3^{(0)}\\\vdots\\x_n^{(0)}\end{bmatrix},\ B=\begin{bmatrix}-z_2^{(1)} & 1\\-z_3^{(1)} & 1\\\vdots & \vdots\\-z_n^{(1)} & 1\end{bmatrix}$$

According to Eq. (4), the solution of $X^{(1)}$ at time $k$ :

$$\hat{x}^{(1)}(k)=(x_1^{(0)}-\frac{\hat{b}}{\hat{a}})\mathrm{e}^{-\hat{a}(k-1)}+\frac{\hat{b}}{\hat{a}},\quad k=1,2,\cdots, \quad (7)$$

Applying the inverse accumulated generating operation (IAGO), the following grey model can be established,

$$\hat{x}^{(0)}(k)=(x_1^{(0)}-\frac{\hat{b}}{\hat{a}})(1-\mathrm{e}^{\hat{a}})(\mathrm{e}^{-\hat{a}(k-1)}),\quad k=2,3,\cdots \quad (8)$$

where $\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\cdots,\hat{x}^{(0)}(n)$ are called the GM(1,1) fitted sequence, while $x^{(0)}(n+1),x^{(0)}(n+1),\cdots$ are called the GM(1,1) forecast values.

As we can see,the GM(1,1) model has the following problem:

1)It need to face the reasonable selection of background value, but there is no unified standard.

2)The initial value in GM(1,1) model take the first data,it has been proved to be unreasonable[7] .

3)The transformation from discrete form to continuous form is lack of rigorous theory basis, therefore, the parameters estimation method can't avoid the jumping errors which come from the difference equation to differential equation.

### 2.2BP neural network

The field of neural networks can be thought of as being related to artificial intelligence, machine learning, parallel processing, statistics, and other fields. The attraction of neural networks is that they are best suited to solving the problems that are the most difficult to solve by traditional computational methods. One algorithm which has hugely contributed to neural network fame is the back-propagation(BP) algorithm. A BP network learns by example, that is, we must provide a learning set that consists of some input examples and the known-correct output for each case. So, we use these input-output examples to show the network what type of behavior is expected, and the BP algorithm allows the network to adapt.

The BP learning process works in small iterative steps, and the network produces some output based on the current state of it's synaptic weights (initially, the output will be random). This output is compared to the known-good output, and a mean-squared error signal is calculated. The error value is then propagated backwards through the network, and small changes are made to the weights in each layer. The weight changes are calculated to reduce the error signal for the case in question. The whole process is repeated for each of the example cases, then back to the first case again, and so on. The cycle is repeated until the overall error value drops below some pre-determined threshold. At this point we say that the network has learned the problem "well enough" - the network will never exactly learn the ideal function, but rather it will asymptotically approach the ideal function. A BP neural network is not the most efficient nor the most accurate way to solve all problem, however, you can use back-propagation for many problems, including a pattern classification problem, compression, prediction and digital signal processing.

### 2.3Modeling basic principle

In order to overcome above-mentioned problems in GM(1,1) model, the GM(1,1) model based on BP neural network is presented.

Assume that

$X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ is a raw sequence, considering the differential equation which satisfies the following initial value problem

$$\begin{cases}\dfrac{dy}{dt}+ay=b\\ y(1)=\hat{x}^{(0)}(1)\end{cases} \tag{9}$$

where $a,b,\hat{x}^{(0)}(1)$ are undetermined parameters.

The above equation can be solved as follows:

$$\frac{dy}{(y(t)-\frac{b}{a})}=-adt$$

$$\ln(y(t)-\frac{b}{a})=-at+c$$

$$y(t)=e^{-at}e^{c}+\frac{b}{a},$$

Since $y(1)=\hat{x}^{(0)}(1)$ thus

$$\hat{x}^{(0)}(1)=e^{-a}e^{c}+\frac{b}{a},\quad e^{c}=e^{a}(\hat{x}^{(0)}(1)-\frac{b}{a}),$$

Therefore, the above differential equation solution contains undetermined parameters is described by

$$y(t)=(\hat{x}^{(0)}(1)-\frac{b}{a})e^{-a(t-1)}+\frac{b}{a}, \tag{10}$$

Let:

$$\begin{aligned}y^{(0)}(t)&=y(t)-y(t-1)\\&=(\hat{x}^{(0)}(1)-\frac{b}{a})(1-e^{a})e^{-a(t-1)}\end{aligned}, \tag{11}$$

Compared with GM(1,1) model, the function value of Eq. (11) at $t$ ( $t=2,3,\cdots,n$ )can be taken as the predicted value in the corresponding time $t$ of GM(1,1) model. In order to obtain a better solution, this initial value does not take the first data of the original sequence data, but as a parameter to be determined by data fitting. Hence, the original discrete data sequence $X^{(0)}$ can be fitted by Eq. (11). The key question is how to estimate the parameters by original discrete data $X^{(0)}$.

**2.3GM(1,1) model and BP neural network mapping relationships**

Based on the general parameter evaluation criteria of function fitting, error function is defined as follows:

$$E=\frac{1}{2n}\sum_{t=1}^{n}(y^{(0)}(t)-x^{(0)}(t))^{2}, \tag{12}$$

Because the error function is nonlinear, we now resolve the above problem by means of BP network which can approximate nonlinear function with any precision[10]. We define $x^{(0)}(t)$ and $y^{(0)}(t)$ as the training example target value and the network output, The error $E$ contains undetermined parameters $a,b,\hat{x}^{(0)}(1)$, so we should establish corresponding relationships of the BP network weight and the parameters. Then the BP algorithm can implement a gradient descent search through the space of possible network weights, iteratively reducing the error $E$ between the training example target values and the network outputs .

Since $e^{-a(t-1)}>0$ , therefore Eq.(11)can be transformed as follows:

$$\begin{aligned}y^{(0)}(t)&=\frac{(\hat{x}^{(0)}(1)-\frac{b}{a})(1-e^{a})e^{-a(t-1)}}{(1+e^{-a(t-1)})}(1+e^{-a(t-1)})\\&=(\hat{x}^{(0)}(1)-\frac{b}{a})(1-e^{a})(1-\frac{1}{1+e^{-a(t-1)}})(1+e^{-a(t-1)})\\&=[(\hat{x}^{(0)}(1)-\frac{b}{a})(1-e^{a})-(\hat{x}^{(0)}(1)-\frac{b}{a})\frac{1}{1+e^{-a(t-1)}}\\&+(\hat{x}^{(0)}(1)-\frac{b}{a})e^{a}\frac{1}{1+e^{-a(t-1)}}](1+e^{-a(t-1)}),\end{aligned} \tag{13}$$

The above equation is mapped to a BP neural network, the network structure is described as Fig.1:

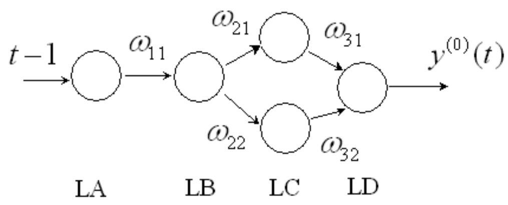


Fig 1: Gray BP neural network structure

The corresponding relationships of the network weight and the parameters are established as follows:

$$\omega_{11}=a,$$

$$\omega_{21}=\frac{b}{a}-\hat{x}^{(0)}(1),$$

$$\omega_{22} = (\hat{x}^{(0)}(1) - \frac{b}{a})e^{a}, \quad (14)$$

$$\omega_{31} = \omega_{32} = 1 + e^{-a(t-1)},$$

The threshold value of $y^{(0)}$ is

$$\theta = (1 + e^{-a(t-1)})(\frac{b}{a} - \hat{x}^{(0)}(1))(1 - e^{a}).$$

And the LB level neuron's transfer function is taken as Sigmoid type functions $f(x) = \frac{1}{1+e^{-x}}$, this function is S-type function, there is a high-gain area which ensure that the network eventually reach a steady[11] , Other neuron's transfer function is taken as a linear function $f(x) = x$.

Through the above method, the GM(1,1) model is mapped to a BP neural network, figure 1 shows this neural network structure for GM(1,1) is simple. The corresponding relationships between GM(1,1) model parameters and BP network weights is established, then GM(1,1) model parameters estimation problem is transformed into the weights of neural network optimization problem. The BP neural network is trained by use of data sets $(t-1, x^{(0)}(t)), t = 1, 2, \cdots, n$.

The BP algorithm learns the weights, it employs gradient descent to minimize the squared error Eq.(12), when the BP network convergence, the optimized model parameters can be extracted, and the optimization modeling for GM(1,1) Model based on BP algorithm can be realized.

As we can see, the GM(1,1) model based on BP neural network doesn't need AGO and IAGO process, compared with the indirect GM(1,1) model, the computation is simpler and the modeling process is more intuitive. On the other hand, it doesn't need transform from discrete form to continuous form and avoid the jumping errors from the difference equation to differential equation, therefore it can overcome the defects in the GM(1,1) model.

## 3. Experiment results and discussions

### 3.1Data sets and experiment design:

In this section, to validate the performance of GM(1,1) model based on BP neural network. The primary data is shown in Table 1.

Table 1:The primary data of simulation experiment

| Sample number | Actual value | Sample number | Actual value |
|---|---|---|---|
| 1 | 2.28 | 6 | 8.84 |
| 2 | 2.98 | 7 | 11.85 |
| 3 | 3.39 | 8 | 12.15 |
| 4 | 4.42 | 9 | 12.71 |
| 5 | 6.86 | | |

The following equation is used to cope with the initial data:

$$\overline{x}^{(0)}(t) = \frac{x^{(0)}(t) - x^{(0)}{}_{\min}}{x^{(0)}{}_{\max} - x^{(0)}{}_{\min}}, \quad (15)$$

where $x^{(0)}(t)$ and $\overline{x}^{(0)}(t)$ are the old and new value, $x^{(0)}{}_{\max}$ and $x^{(0)}{}_{\min}$ are the minimum and maximum value of the original data. Experiments show that the pretreatment of the sample can speed up the network training speed and improve the prediction accuracy. According to the proposed method, the BP neural network with three hidden layer is established. The network structure is described as Fig.1. The first level neuron's transfer function is taken Log-Sigmoid type function, other neuron's transfer function is taken as a linear function. The maximum training epoch is $M = 500$, the allowance permissible error is $E = 0.0001$. The LM (Levenberg -Marquardt) is taken as the training algorithm, the learning rate is dynamically determined by LM algorithm. Using the Matlab6.5 programming, when the network achieves the accuracy requirement, we can obtain the optimized gray BP neural network model.

Figure 2 shows the training times of this method only use 11 steps to reach the minimum error, however, the traditional BP network not only need a large amount of data to train, but also the theoretical guidance of network settings is lack, on the other hand the training time is much greater than this article method. Many experiments show that the number of network convergence the method is generally is about 15. Experiment shows the method combines the characteristics of the small sample and poor information of the GM (1,1) model, and displayed the characteristics of strong nonlinear approximation and fault-tolerant capability of the neural network.

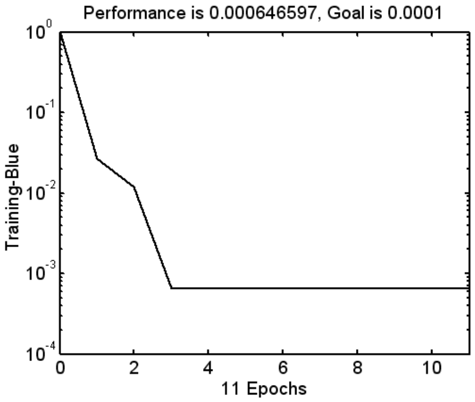


Fig.2 Training errors of the primitive data based on the method in this paper

**3.2 Model validation criteria:**

The results obtained by the model in this paper are compared with traditional GM(1,1) model and other model.

Various models are formed in this paper:

- Model2:Traditional GM(1,1) model.
- Model3: Method in the paper [9]
- Model4: Method in the paper [11].
- Model5: Method in this paper.

Mean absolute percentage error (MAPE) approach[12,13] has been recommended to validation. Mean absolute percentage error is defined as

$$\text{MAPE} = \frac{1}{l}\sum_{i=1}^{l}\left|\frac{y_i - \hat{y}_i}{y_i}\right| , \qquad (16)$$

where $y_i$ is the actual value, and $\hat{y}_i$ is the predicted value, $\left|\frac{y_i - \hat{y}_i}{y_i}\right|$ is absolute percentage error of $y_i$.The model yielded plausible prediction values when MAPE is low, vice versa [13].

**3.3Empirical result and error analysis:**

Based on above-mentioned research frame, this study used 4 models for the experiments. The result of the identification is shown in Table 2. As shown from the results, the identification values in Model5 fit the actual values very well. We can find that the model not only has the nonlinear mapping capacity, but also reserves the features of grey system model, therefore the model for small sample fitting is applicable. Additionally, in view of Figure 2, the network computation speed is more rapid than other neural network, because the gray BP neural network structure is directly built based on the inverse accumulated form of the GM(1,1) model, the structure is more simpler.

Compared with the traditional GM(1,1) model, the precision in this article enhances 84.33%. As shown from the results, the initial value does not take the first data of the original sequence data, but as a parameter to be determined by data fitting. It approximately equals to 2.280, however, the initial value in GM(1,1) model take the first data, it has been proved to be unreasonable.Compared with other existing direct modeling method in the paper[9] and the paper[11], the precision in this article respectively enhances 75.60% and 25.28%. Additionally, these in the paper[9] and the paper[11], the modeling process need to use accumulated generating operation and inverse accumulated generating operation, the computation is complex. Simulation results show that Model5 is feasible and effective. Furthermore, according the criteria purposed by Lewis[13], the model is better than the other forecasting models because of lower average MAPE.

Table2: Forecasted values and MAPE

| Actual value | Model2 | | Model3 | | Model4 | | Model5 | |
|---|---|---|---|---|---|---|---|---|
| | Model value | APE(%) | Model value | APE (%) | Model value | APE % | Model value | APE % |
| 2.28 | 2.28 | 0 | ≈2.280 | ≈0 | 2.280 | 0 | ≈2.280 | ≈0 |
| 2.98 | 3.72 | 24.870 | ≈2.980 | ≈0 | 3.01 | 1.04 | ≈2.980 | ≈0 |
| 3.39 | 4.53 | 33.537 | 3.767 | 11.107 | 3.28 | 3.13 | 3.42 | 0.921 |
| 4.42 | 5.51 | 29.885 | 4.760 | 7.707 | 4.60 | 4.10 | 4.47 | 1.232 |
| 6.86 | 6.70 | 2.338 | 6.017 | 12.285 | 6.62 | 3.54 | 6.54 | 4.673 |
| 8.64 | 8.15 | 5.668 | 7.605 | 11.974 | 9.03 | 4.55 | 9.21 | 6.590 |
| 11.85 | 9.92 | 16.328 | 9.612 | 18.879 | 11.30 | 4.68 | 11.28 | 4.823 |
| 12.15 | 12.06 | 0.723 | ≈12.15 | ≈0 | 12.64 | 3.99 | 12.34 | 1.530 |
| 12.71 | 14.67 | 15.452 | 15.356 | 20.825 | 12.98 | 2.11 | 12.77 | 0.412 |
| MAPE (%) | | 14.311 | | 9.197 | | 2.998 | | 2.242 |

## 4. Conclusion

The practice proves that the traditional and the improved GM (1, 1) modeling method exist some shortage. In this paper, a novel GM(1,1) model based on BP neural network is proposed and manifests good performance in the preliminary experiment. From the data fitting's viewpoint, the GM(1,1) model is mapped to a BP neural network, the corresponding relation between GM(1,1) model parameters and BP network weights is established, GM(1,1) model parameters estimation problem is transformed into the weights of neural network optimization problem. This method dose not need to use AGO and IAGO, which avoids and overcomes the theory defects of the traditional GM (1,1) model. BP neural network is a highly effective function approximation method in practice. The experiment proves that fitting GM(1,1) model by means of the BP network may display GM(1,1) model and the neural network modeling merit, and enhance the forecast precision. However, as a BP neural network, the GM(1,1) model based on BP neural network is easy to converge toward some local minimum. From this angle, further research work should be studied.

## References


[1]Deng Julong. Contral problems of grey system[J]. Systems& Contral Letters. 1982, 1:288- 294.

[2] Deng Julong. Introduction to grey system theory[J]. The Journal of Grey system.1989, 1: 1- 2.

[3] Kayacan E, Ulutas B, Kaynak O. Grey system theory-based models in time series prediction[J]. Expert Systems with Applications, 2010, 37:1784-1789.

[4] Song Qiang, Wang Aiming. Simulation and Prediction of Alkalinity in Sintering ProcessBased on Grey Least Squares Support Vector Machine[J].Journal of Iron and Steel Research, International, 2009, 16(5): 1-6.

[5] Zhou Deqiang.GM（1,1）model based on least absolute deviation and application in the power load forecasting[J]. Power System Protection and Control, 2011, 39(1):100-103 (in Chinese).

[6] Zeng Bo, Liu Sifeng and Xie Naiming. Prediction model of interval grey number based on DGM(1,1)[J].Journal of Systems Engineering and Electronics. 2010, 21(4):598-603,

[7] Zhang Dahai, Jiang Shifang, Shi Kaiquan. Theoretical Defect of Grey Prediction formula and Its improvement[J].SystemsEngineering-Theory&Practice, 2002, 22(8):140-142 (in Chinese).

[8] Xie Naiming, and Liu Sifeng. Discrete GM(1, 1) and mechanism of grey forecasting model[J]. Systems Engineering-Theory & Practice, 2005,25(1): 93–99 (in Chinese).

[9] Zheng Zhaoning, and Liu De Shun. Direct Modeling Improved GM (1, 1) Model IGM (1, 1) by Genetic Algorithm[J].Systems Engineering Theory& Practice, 2003, 23(5):99-102 (in Chinese).

[10] Shi Weiren, Wang Yanxia, Tang Yunjian, et al .Water quality parameter forecast based on grey neural network modeling[J].Journal of Computer Applications, 2009, 29(6): 1529-1531,1535 (in Chinese).

[11]Zhong Lu, Bai Zhengang, Xia Hongxia, et al. Optimization and Application of Neural Network Modeling for Gray Problem[J]. Computer Engineering And Applications, 2001, 37(9): 33-35 (in Chinese).

[12] Hsu Lichang, Wang Chaohung. Forecast the output of integrated circuit industry using a grey model improved by the Bayesian analysis[J]. Technological Forecasting and Social Change, 2007, 74(6): 843-853.

[13]Lewis,C. Industrial and Business Forecasting Methods[M]. Butterworth Scientific, London. 1982.


**Deqiang Zhou:** Associate professor of the school of information and mathematics in Yangtze University, interested in machine learning theory, grey systems theory and complicated mathematical models.