

Time Series Forecasting Model Based on Discrete Grey LS-SVM

De-qiang Zhou

School of Information and Mathematics, Yangtze University, Jing zhou, 434023, China
Email: zdqmfk@163.com

Abstract—The advantages and disadvantages of discrete GM(1,1) model and least squares support vector machine are analyzed respectively, this article proposes a new time series forecasting model of discrete grey least squares support vector machine. The new model adopts structural risk minimization principle, at the same time develops the advantages of accumulation generation in the grey forecasting method, weakens the effect of stochastic-disturbing factors in original sequence, and avoids the theoretical defects existing in the grey forecasting model. The simulation results show that the forecasting model is effective and reliable, and consolidates the advantage of the discrete GM(1,1) model and least squares support vector machine. It offers a new way to improve the time series forecasting accuracy.

Index Terms—Time series prediction, Least square support vector machines algorithm, Grey system, Small samples, Discrete GM(1,1) model, Discrete grey least squares support vector machine .

I. INTRODUCTION

Time series prediction refers to the process by which the future values of a system is forecasted based on the information obtained from the past and current data points[1]. At present, there are a lot of methods for time series prediction, from traditional statistical method such as Auto Regressive Moving Average model to artificial intelligence based approaches. Grey system theory is an interdisciplinary scientific area that was first introduced in early 1980s by Deng[2]. Since then, the theory has become quite popular with its ability to deal with the systems that have partially unknown parameters. 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[3]. Because the grey prediction model needs little origin data, has simple calculate process and higher forecasting accuracy, it has been widely used in the time series prediction of a lot of research fields.

The GM(1,1) model is widely discussed and studied in the grey system[2,3]. The purpose of GM (1, 1) model is to work on system forecasting with poor, incomplete or uncertain messages. The GM (1,1) is based on the grey generating function, which uses the variation within the system to find the relations between sequential data and then establish the prediction mode. However, many scholars find that there are many theory defects in traditional GM (1,1) model[4], and do a lot of researches for this[5,6,7,8]. Discrete GM(1,1) (DGM(1,1))model is a

new grey prediction model, which improve some problem existent in original GM(1,1)[7]. However, the parameters estimation method of the DGM(1,1) is still by means of least square method. Least square method is based on empirical risk minimization (ERM) principal, which is large sample principle[9]. Furthermore, the GM(1,1) forecasting study is small sample prediction, estimating parameters based on least square method does not match with the small sample modeling concept. On the other hand, the robustness of the least square method is poor, which affects the stability of the DGM(1,1) model[6,10]. To sum up the above reasons, it is not suitable to estimate DGM(1,1) model parameters by means of least square method.

According to the above describing, we need find a new method to improve the precision of DGM(1,1) model. Recently, a new intelligent approach support vector machine (SVM) has gained popularity due to its many attractive, analytic and computational features, and promising empirical performance[9,11,12]. It is a promising technique owing to its successful application in classification and regression tasks[13,14]. SVM comprises an important class of learning algorithms based on statistical learning. Moreover, SVM is established based on the structural risk minimization(SRM) principle. It shows the good generalization ability. Least squares support vector machine (LS-SVM) has received considerable attention for the last few years due to its simplicity in problem formulation and implementation. It was first proposed by Suykens and Vandewalle[13]. LS-SVM is a least square version for SVM that involves equality instead of inequality constraints and works with a least squares cost function, which leads to the advantages of fast convergence, high accuracy and low computational efforts[11,13]. LS-SVM has been successfully applied to real application modeling considered as a non-linear function regression problem.

For a time series prediction problem, the core of these models is to establish a prediction model[2]. However, in the actual forecast, the same problem can have many different forecasting methods. Given many approximation methods, it has become difficult for an engineer to select the most appropriate method for the problems under study[15]. So it would be very meaningful if different methods can be combined, which consolidates the advantage of the different models and obtains better forecasting accuracy. Based on the approach of “mixture of experts”, some researchers have

integrated the GM(1,1) and the artificial neural network to enhance the prediction according to their complementary merits[16,17]. The moving average autoregressive exogenous prediction model is combined with grey predictors for time series prediction[18], and it is proved that the hybrid method has a greater forecasting accuracy than the GM(1,1) method. Another study introduces a support vector regression grey model (SVRGM) which combines support vector regression (SVR) learning algorithm and grey system theory to obtain a better approach to time series prediction[19].

Both the grey model and the SVM model do not need to know whether the prediction variables obey normal distribution, do not to require too much statistic sample. However, each has some defects. The primary interest of this paper lies in combining the two models: DGM(1,1) forecasting model and LS-SVM forecasting model. Based on the theoretical analysis, a new idea or algorithm is put forward. This paper establishes a new DGM(1,1) time series prediction model based on LS-SVM algorithm, which is a combination of DG (1,1) and LS-SVM with their respective advantages, grey theory is used to conduct accumulative sequence of the raw data, and the least squares support vector machine is adopted for the SRM and prediction.

The remainder of this paper is organized as follows: Sec.2 are brief overviews of DGM(1,1) and LS-SVM. Then a time series prediction model based on discrete grey least squares support vector machine is established in Section 3. Experimental results are given in Section 4. Finally, conclusion is drawn in Section 5.

II. REVIEW ON DGM(1,1) MODEL AND LS-SVM

A. DGM(1,1) model

Definition1: Let $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ is a sequence, where $x^{(0)}(k) \geq 0, k = 1, 2, \dots, n$ and $X^{(1)}$ is the accumulated generating operation (1-AGO) sequence of $X^{(0)}$, $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$,

Where

$$x^{(1)}(k) = \sum_{j=1}^k x^{(0)}(j), (k = 1, 2, \dots, n)$$

$$x^{(1)}(k+1) = \beta_1 x^{(1)}(k) + \beta_2 \quad (1)$$

Then (1) is called a DGM(1,1) model, or the discrete form of a GM(1,1) model[7].

Theorem1: If $\hat{\beta} = (\beta_1, \beta_2)^T$ is a sequence parameters and

$$B = \begin{bmatrix} x^{(1)}(1) & 1 \\ x^{(1)}(2) & 1 \\ \vdots & \vdots \\ x^{(1)}(n-1) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(1)}(2) \\ x^{(1)}(3) \\ \vdots \\ x^{(1)}(n) \end{bmatrix} \quad (2)$$

Then the least square estimate sequence of the grey differential (1) satisfies

$$\hat{\beta} = (B^T B)^{-1} B^T Y.$$

Theorem2: If B , Y and $\hat{\beta}$ are stated in theorem1, then let

$$x^{(1)}(1) = x^{(0)}(1), k = 1, 2, \dots, n-1,$$

then

$$\hat{x}^{(1)}(k+1) = \beta_1^k x^{(0)}(1) + \frac{1-\beta_1^k}{1-\beta_1} \beta_2 \quad (3)$$

The restored values, $k = 1, 2, \dots, n-1$, then

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (4)$$

The traditional DGM(1,1) model has the following problem:

(1)The DGM(1,1) estimates parameters based on least square method, which does not match with the small sample modeling of GM(1,1).

(2)The robustness of the least square method is poor, which affects the stability of the model.

B. LS-SVM algorithm

Giving a training set $\{(x_k, y_k)\}_{k=1}^N, x_k \in \mathbf{R}^n, y_k \in \mathbf{R}$,

where x_k is the k -th input sample, y_k is the desired output value. LS-SVM has been developed for solving an approximation problem. It allows the construction of an approximation by mapping the initial data set implicitly into some high dimensional feature space Z through some mapping $\varphi(x)$. LS-SVM aims at constructing a simple linear approximation in the feature space Z then corresponds to a nonlinear approximation in sample space[20].

All can be done implicitly in Z by means of the kernel function $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$. LS-SVM models take the form:

$$f(x) = w^T \varphi(x) + b \quad (5)$$

where $w \in \mathbf{R}^k$ is weight vector and $b \in \mathbf{R}$ is bias term. To obtain w and b one solves the following constrained optimization problem[13]:

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + \frac{\lambda}{2} \sum_{k=1}^N \xi_k^2 \quad (6)$$

$$\text{s.t. } y_k = w^T \varphi(x_k) + b + \xi_k, k = 1, 2, \dots, N$$

where ξ_k is a error variable, $\lambda > 0$ is tradeoff parameter between a smoother solution and training errors.

To solve the optimization problem, the Lagrange function can be defined as follows:

$$L(w, b, \xi, \alpha) = \frac{1}{2} \|w\|^2 + \frac{\lambda}{2} \sum_{k=1}^N \xi_k^2 - \sum_{k=1}^N \alpha_k [w^T \varphi(x_k) + b + \xi_k - y_k] \quad (7)$$

where α_k are Lagrange multipliers. The solution of \mathbf{a} and b is given by

$$\begin{bmatrix} 0 & \mathbf{e}^T \\ \mathbf{e} & \Omega + \frac{\mathbf{I}}{\lambda} \end{bmatrix}_{N \times N} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (8)$$

where, $\mathbf{e} = (1, 1, \dots, 1)_{1 \times N}^T$, $\mathbf{I} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \dots & \vdots \\ 1 & \dots & 1 \end{pmatrix}_{N \times N}$

and $\Omega = (k(x_i, x_j))_{N \times N}$.

Finally, we substitute w into Eq. (1) and thus the training sets of nonlinear approximation is obtained too,

$$f(x) = \sum_{k=1}^N \alpha_k K(x_k, x) + b \quad (9)$$

The selection of the kernel function has several possibilities. It is arbitrary symmetric function, which meets the Mercer theorem[9], such as:

- (1) linear Kernel: $K(x_i, x_j) = x_i^T x_j$
- (2) polynomial Kernel of degree d :

$$K(x_i, x_j) = (x_i^T x_j + c)^d, c > 0$$

Where c is a parameter.

- (3) Kernels generating Gaussian radial basis function(RBF):

$$K(x_i, x_j) = \exp\{-\|x_i - x_j\|^2 / 2\sigma^2\}$$

For this equation, σ is a positive real constant which can be adjusted by users. Because Gaussian RBF kernels tend to give good performance under general smoothness assumptions, since Gaussian RBF function is usually used as a kernel function. The precision and convergence of LS-SVM are both affected by λ and σ . The selection of parameters of a SVM model is important to the accuracy of the forecasting. Most SVM practitioners select these parameters empirically by trying a finite number of values and keeping those that provide the least testing error[21].

Unfortunately, now there is no unified method to select the parameters of the model.

III. DISCRETE GREY LEAST SQUARES SUPPORT VECTOR MACHINE

Against defects in a single model described above, a new discrete GM(1,1) time series prediction model based on LS-SVM algorithm is established.

Although grey forecast and LS-SVM forecast build on the basis of different theories, there are some similarities between DGM(1, 1) and LS-SVM. Both of their study is small sample prediction. DGM(1,1) by identifying the parameters of the model is actually based on the least squares linear regression, whereas support vector machine is evolved from the linear optimal surface. Both models have their own advantages and weaknesses. DGM(1,1) is a model of the difference equation, which strengthens the regularity of raw data by cumulative generation; moreover, it is the fitting of exponential curve. However, the DGM(1,1) estimates parameters by means of least square method[7], it requires too much statistic sample to ensure the ability of the model, and it need to require the prediction variables obey normal distribution, despite the grey model itself does not make demands on data distribution. On the other hand, the robustness of the least square method is poor, which affects the stability of the model. LS-SVM is a theory based on SRM, which has very good generalization ability[19]. LS-SVM method does not need to require the prediction variables obey normal distribution.

Based on the above analysis, combining the two above mentioned small sample prediction methods, a new grey LS-SVM model named DGM(1,1) LS-SVM is proposed to overcome these limitations on both of the two forecasting models. The fluctuation of data sequences is weakens by the grey theory, and the generalization ability of the model is enhanced by the statistical learning theory, and the DGM(1,1) LS-SVM is a combination of those advantages.

Above all, the DGM(1,1) is used to conduct a cumulative sequences of the raw data, and is employed to choose the kernel of the LS-SVM, on the other hand, the LS-SVM algorithm is adopted for the process and prediction.

Theorem3: If $(\beta_2, \beta_1)^T$ is the grey differential (1) parameters and

$$B = \begin{bmatrix} 0 & \mathbf{e}^T \\ \mathbf{e} & \phi(x^{(1)})\phi(x^{(1)})^T + \frac{\mathbf{I}}{\lambda} \end{bmatrix}, Y = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}$$

where $\mathbf{e} = (1, 1, \dots, 1)_{1 \times (n-1)}^T$, $\lambda > 0$,

$$\mathbf{I} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \dots & \vdots \\ 1 & \dots & 1 \end{pmatrix}_{(n-2) \times (n-2)},$$

$$\mathbf{y} = [x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)]^T,$$

and $\phi(x^{(1)}) = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n-1))^T$. Then the LS-SVM algorithm estimate sequence of the grey difference (1) satisfies

$$\begin{pmatrix} \beta_2 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ 0 & \phi(x^{(1)})^T \end{pmatrix}_{2 \times n} (B^T B)^{-1} B^T Y$$

Proof: To consider

$$x^{(1)}(k+1) = \beta_1 x^{(1)}(k) + \beta_2, \quad k=1,2,\dots,n-1, \quad (10)$$

taking $\{(x^{(1)}(k), x^{(1)}(k+1))\}_{k=1}^{n-1}$ as a training set,

coefficients $(\beta_2, \beta_1)^T$ can be obtained as solution to the following optimization problem based on LS-SVM algorithm,

$$\begin{aligned} \min_{\beta_1, \beta_2, \xi} & \frac{1}{2} \|\beta_1\|^2 + \frac{\lambda}{2} \sum_{k=1}^{n-1} \xi_k^2 \\ \text{s.t.} & x^{(1)}(k+1) = \beta_1 x^{(1)}(k) + \beta_2 + \xi_k, \\ & k=1,2,\dots,n-1 \end{aligned} \quad (11)$$

To solve the optimization problem, one defines the Lagrange function

$$\begin{aligned} L(\beta_1, \beta_2, \xi, \alpha) &= \frac{1}{2} \|\beta_1\|^2 + \frac{\lambda}{2} \sum_{k=1}^{n-1} \xi_k^2 \\ &- \sum_{k=1}^{n-1} \alpha_k [\beta_1 x^{(1)}(k) + \beta_2 + \xi_k - x^{(1)}(k+1)] \end{aligned} \quad (12)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_{n-1}]^T$ are Lagrange multipliers.

By Karush-kuhn-Tucker conditions, the conditions for optimality are[13]

$$\begin{cases} \frac{\partial L}{\partial \beta_1} = 0 \rightarrow \beta_1 = \sum_{k=1}^{n-1} \alpha_k x^{(1)}(k), \\ \frac{\partial L}{\partial \beta_2} = 0 \rightarrow \sum_{k=1}^{n-1} \alpha_k = 0, \\ \frac{\partial L}{\partial \xi_k} = 0 \rightarrow \alpha_k = \lambda \xi_k \\ \frac{\partial L}{\partial \alpha_k} = 0 \rightarrow \beta_1 x^{(1)}(k) + \beta_2 + \xi_k - x^{(1)}(k+1) = 0 \end{cases}$$

where $k=1,2,\dots,n-1$.

After elimination of β_1 and ξ_k , the solution is given by

$$\begin{bmatrix} 0 & \mathbf{e}^T \\ \mathbf{e} & \phi(x^{(1)})\phi(x^{(1)})^T + \frac{\mathbf{I}}{\lambda} \end{bmatrix}_{n \times n} \begin{bmatrix} \beta_2 \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (13)$$

where $\mathbf{e} = (1, 1, \dots, 1)_{1 \times (n-1)}^T$, $\mathbf{I} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \dots & \vdots \\ 1 & \dots & 1 \end{pmatrix}_{(n-1) \times (n-1)}$,

$$\alpha = [\alpha_1, \alpha_2, \dots, \alpha_{n-1}]^T, \quad \mathbf{y} = [x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)]^T$$

$$\phi(x^{(1)}) = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n-1))^T.$$

Let $B = \begin{bmatrix} 0 & \mathbf{e}^T \\ \mathbf{e} & \phi(x^{(1)})\phi(x^{(1)})^T + \frac{\mathbf{I}}{\lambda} \end{bmatrix}$, $Y = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}$, then

$$\begin{pmatrix} \beta_2 \\ \alpha \end{pmatrix}_{n \times 1} = (B^T B)^{-1} B^T Y, \text{ according to}$$

$$\phi(x^{(1)}) = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n-1))^T,$$

$$\beta_1 = \sum_{k=1}^{n-1} \alpha_k x^{(1)}(k) = \phi(x^{(1)})^T \alpha \quad (14)$$

hence

$$\begin{aligned} \begin{pmatrix} \beta_2 \\ \beta_1 \end{pmatrix} &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \phi(x^{(1)})^T \end{pmatrix}_{2 \times n} \begin{pmatrix} \beta_2 \\ \alpha \end{pmatrix}_{n \times 1} \\ &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \phi(x^{(1)})^T \end{pmatrix}_{2 \times n} (B^T B)^{-1} B^T Y \end{aligned} \quad (15)$$

DGM(1,1) model can be expressed

$$\hat{x}^{(1)}(k+1) = \sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) x^{(1)}(k) + \beta_2 \quad (16)$$

Let $K(x^{(1)}(i), x^{(1)}(k)) = x^{(1)}(i) x^{(1)}(k)$, where the function K is linear Kernel Kernel function. Formula (16) can be expressed

$$\hat{x}^{(1)}(k+1) = \sum_{i=1}^{n-1} \alpha_i K(x^{(1)}(i), x^{(1)}(k)) + \beta_2 \quad (17)$$

(17) is a LS-SVM model based on DGM(1,1), we call it a DGM(1,1) LS-SVM model.

Let $x^{(0)}(1) = x^{(1)}(1)$, then restored values,

$$\begin{aligned} \hat{x}^{(0)}(k+1) &= \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = \sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) x^{(0)}(k) \\ &k=1,2,\dots,n-1. \end{aligned} \quad (18)$$

Based on linear Kernel function, we can get a new method: DGM(1,1) based on LS-SVM(Linear Support Vector Machine, LS-SVM). The LS-SVM training procedure involves a section of the kernel parameter and the tradeoff parameter. From (17), we can see that we only use linear Kernel to represent the Kernel function of DGM(1,1) LS-SVM. There is only one parameter λ need to be made certain, and the number of parameters of this LS-SVM is smaller than other kind of SVM. On the other hand, the AGO functions can transfer the disordering series into monotone raising ones, which provided LS-SVM learning with reasonable data base, the uncertain factors are decreased. For DGM(1,1) modeling based on LS-SVM, the fluctuation of data sequence is weakened by the grey theory and the support vector machine is adopted for the SRM, so the LS-SVM DGM(1,1) machine is a combination of those advantages[5].

In fact, formula (18) gives the DGM(1,1) LS-SVM model expression for fitting. We now discuss the forecasting problem of a DGM(1,1) LS-SVM model.

(1) If $k = n$, then

$$\hat{x}^{(1)}(n+1) = \sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) x^{(1)}(n) + \beta_2, \text{ therefore,}$$

$$\hat{x}^{(0)}(n+1) = \hat{x}^{(1)}(n+1) - x^{(1)}(n)$$

$$= \left(\sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) - 1 \right) x^{(1)}(n) + \beta_2 \quad (19)$$

(2) If $k = n+1, n+2, \dots$, because $x^{(0)}(k)$ is unknown, therefore $x^{(1)}(n)$ is also unknown, now based on iteration

$$\hat{x}^{(1)}(k+1) = \sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) \hat{x}^{(1)}(k) + \beta_2,$$

forecasting value can be described as follow:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$$

$$= \sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) (\hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1)) \quad (20)$$

$$= \sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) \hat{x}^{(0)}(k)$$

Combining (19) and (20), one arrives at

$$\begin{cases} \hat{x}^{(0)}(k+1) = \sum_{i=1}^{n-1} \alpha_i x^{(1)}(i) \hat{x}^{(0)}(k), \\ k = n+1, n+2, \dots \\ \hat{x}^{(0)}(n+1) = \hat{x}^{(1)}(n+1) - x^{(1)}(n) \end{cases} \quad (21)$$

the later development trend of raw series $X^{(0)}$ can be forecasted.

IV. CALCULATION EXAMPLE ANALYSES

In order to test the accuracies of the models in this paper, we now use the following example to carry on the experiment. The raw series [8] demonstrate as follows:

$$X^{(0)} = (2.28, 2.98, 3.39, 4.42, 6.86, 8.64, 11.85, 12.15, 12.71)$$

The first six data are taken as the original training samples, the latter 3 data as test samples. The results obtained by the models in this paper are compared with traditional GM(1,1) model and DGM(1,1) model.

Various models are formed in this paper:

- GM(1,1): GM(1,1) model.
- DGM(1,1): Discrete GM(1,1) model.
- LDGM(1,1): DGM(1,1) LS-SVM.

The results of the identification can be described as Figure 1-3. The “*” expresses the actual value, the real line is the model value. Figure 1 is the identification which uses the GM(1,1) model. Figure 2 is the identification which uses the DGM(1,1) model. Figure 3 is the identification, which uses the DGM (1,1) LS-SVM model.

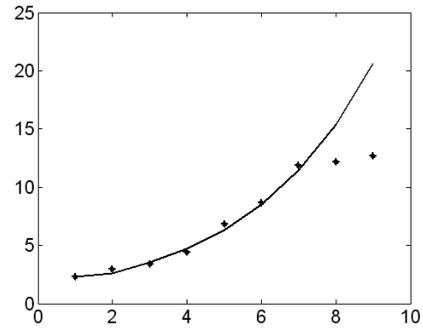


Fig. 1. The identification based on GM(1,1)

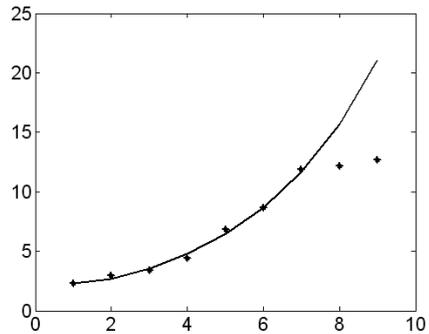


Fig. 2. The identification based on DGM(1,1)

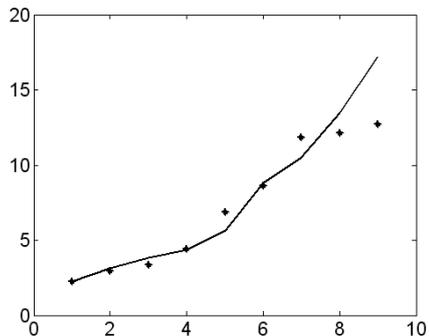


Fig. 3. The identification based on LDGM(1,1)

To demonstrate the accuracy of above models, the actual value and the predicted value can be compared. Eqs.(22)-(24) are the three accuracy evaluation standards that are used to examine the accuracy of the models in this study.

1) The average absolute error:

$$MAE = \frac{1}{l} \sum_{i=1}^l |y_i - \hat{y}_i| \quad (22)$$

2) The average absolute relative percentage error:

$$MA R P E = \frac{1}{l} \sum_{i=1}^l \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (23)$$

3) The mean square relative error:

$$MSR E = \sqrt{\frac{1}{l} \sum_{i=1}^l \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2} \quad (24)$$

where y_i is the actual value, and \hat{y}_i is the predicted value.

The result of the identification can be described as Table 1-2. Table 1 shows the identification index on training samples of different methods.

Table 1. Index on training samples of different methods being compared

Index	GM(1,1)	DGM(1,1)	LDGM(1,1)
MAE	0.2366	0.2208	0.3345
MARPE(%)	5.2684	5.1859	6.4692
MSRE(%)	6.6529	6.5368	9.1104

Table 2 shows the identification index on test samples of different methods

Table 2. Index on test samples of different methods being compared

Index	GM(1,1)	DGM(1,1)	LDGM(1,1)
MAE	1.9292	2.0150	1.1956
MARPE(%)	15.411	16.085	9.613
MSRE(%)	27.769	29.395	15.873

Figure 1-3, show that the training samples are approximate to exponential distribution. Table 1 show that the fitting effect of GM(1,1) and DGM(1,1) model is good. That is to say, the precision, with applying the GM(1,1) and DGM(1,1) model to forecast an approximate exponential incremental sequence is satisfactory. However, it can also be noticed in Figure 1-3, in test samples, nonlinear feature is strong, for comparison, the forecasting performance of GM(1,1) and DGM(1,1) model is inferior to LDGM(1,1). The result of the identification is listed in Table 2.

Based on LS-SVM algorithm, LDGM(1,1) find a balance between a smoother solution and training errors, thus the forecasting accuracy of this model in this paper is higher than that of a single GM (1, 1) model or the DGM (1,1). Figure 3 shows that the generalization performance of LDGM(1,1) is improved. LDGM(1,1) is capable of processing nonlinear adaptable information. Table 2 shows that the better generalization performance and forecasting accuracy are achieved in LDGM(1,1). The results demonstrate that the prediction accuracy of new model are superior to a single model, which exhibits excellent learning ability with fewer training data and the generalization capability.

In summary, it can be concluded that LDGM(1,1) model shows excellent learning ability with fewer training data, and combines the advantages of DGM(1, 1) and LS-SVM.

V. CONCLUSION

A new discrete grey least squares support vector machine is proposed in this paper as a method for time series prediction. This algorithm combines the advantages of DGM(1, 1) and LS-SVM. Through an example's program simulation, the validity and practicality of this model is verified. Although the proposed model

overcomes the defects of parameters estimation in traditional DGM(1,1) model, they may be superior to other modeling methods in some aspects. There are some potential drawbacks such as the selection of tradeoff parameter related to a least squares cost function. The performance of this model is more related to the selection of trade-off parameter. Thus, how to quickly and accurately select the model trade-off parameter should be further studied.

ACKNOWLEDGMENT

This paper is supported by the Hubei province Nature Science Fund of China(2013CFA053) and Open Research Fund Program of Institute of Applied Mathematics Yangtze University of China(KF1506).

REFERENCES

- [1] Kayacan. E, Ulutas. B, & Kaynak. O, "Grey system theory-based models in time series prediction", *Expert Systems with Applications*, Vol. 33, No. 2, pp. 1784-1789, Mar. 2010.
- [2] Julong, Deng., "Contra problems of grey system", *Systems & Control Letters*, Vol. 1, No. 5, pp. 288- 294, Mar. 1982.
- [3] Bo Zeng, Sifeng Liu, and Naiming Xie, "Prediction model of interval grey number based on DGM(1,1)", *Journal of Systems Engineering and Electronics*, Vol. 21, No. 4, pp. 598-603, Aug. 2010.
- [4] Dahai Zhang, Sifang Wang, and KaiQuang Shi, "Theoretical Defect of Grey Prediction Formula and Its Improvement", *Systems Engineering-Theory & Practice*, Vol. 22, No. 8, pp. 140-142, Aug. 2002.
- [5] GuanJun Tan, "The Structure Method and Application of Background Value in Grey System GM(1,1) Model(I)", *Systems Engineering-Theory & Practice*, Vol. 20, No. 4, pp. 98-103, Apr. 2000.
- [6] Deqiang Zhou, "GM (1,1) model based on least absolute deviation and application in the power load forecasting", *Power System Protection and Control*, Vol. 39, No. 1, pp. 100-103, Jan. 2011.
- [7] Naiming Xie, and Sifeng Liu, "Discrete GM(1, 1) and mechanism of grey forecasting model", *Systems Engineering-Theory & Practice*, Vol. 25, No. 1, pp. 93-99, Jan. 2005.
- [8] Zhaoning Zheng, Deshun Liu, "Direct Modeling Improved GM (1, 1) Model IGM (1, 1) by Genetic Algorithm", *Systems Engineering Theory & Practice*, Vol. 23, No. 5, pp. 99-102, May, 2003.
- [9] Vapnik. V. N, *The nature of statistical learning theory*, Springer -Verlag, New York, 1995.
- [10] Dehong An, Wenxiu Han, and Yihong Yue, "Improved combination forecast method and its application in short-term load forecasting of a power system", *Journal of Systems Engineering and Electronics*, Vol. 26, No. 6, pp. 842-844, Jun. 2006.
- [11] Hung. Y.H, and Liao. Y.S, "Application PCA and Fixed Size LS-SVM Method for Large Scale Classification Problems", *Information Technology Journal*, Vol. 7, No. 6, pp. 890-896, Jun. 2008.
- [12] Wu. F.F, and Zhao. Y.L, "Least Square Support Vector Machine on Molet Wavelet Kernel Function and its Application to Nonlinear System Identification",

Information Technology Journal, Vol.5,No.3,pp.439-444, Mar.2006.

- [13] Suykens. J. A. K, and Vandewalle. J, “Least squares support vector machine classifiers”, *Neural Processing Letters*, Vol.9, No.3,pp.293-300, Jun.1999.
- [14] Wang.L. J,Lai.H. C,and Zhang.T. Y, “An Improved on Least Square Support Vector Machines”, *Information Technology Journal*, Vol.7,N0.2, pp.370-373, Feb.2008.
- [15] Yatong Zhou, Taiyi Zhang, and Liejun Wang, “On the Relationship between LS-SVM, MSA, and LSA”, *International Journal of Computer Science and Network Security*, Vol.6 No.11, pp. 01-05, Nov.2006.
- [16] C. C. Chiang, M. C. Ho, and J. A. Chen. “A hybrid approach of neural networks and grey modeling for adaptive electricity load forecasting”, *Neural Computing & Applications*, Vol.15,No.3,pp.328–338, Mar,2006.
- [17] John Paul T. Yusiong, “Optimizing Artificial Neural Networks using Cat Swarm Optimization Algorithm”, *International Journal of Intelligent Systems and Applications*, Vol.5,No.1,pp. 69-80, Dec. 2012.
- [18] K. Y., Huang, C. J. Jane. “A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories”, *Expert Systems with Applications*, Vol.36,No.3,pp. 5387-5392, 2009.
- [19] B. R. Chang, H. F. Tsai. “Forecast approach using neural network adaptation to support vector regression grey model and generalized auto-regressive conditional heteroscedasticity”, *Expert Systems with Applications*, Vol.34,pp.925–934, 2008.
- [20] Qiang Song, and Ai-min Wang, “Simulation and Prediction of Alkalinity in Sintering ProcessBased on Grey Least Squares Support Vector Machine”, *Journal of Iron and Steel Research, International*, Vol.16, No.5,pp.01-06, Sept.2009.
- [21] Xuemei Li, Ming Shao,and Lixing Ding, “Particle Swarm Optimization-based LS-SVM for Building Cooling Load Prediction”, *Journal of Computers*, Vol. 5, No. 4, pp.614-621, Apr. 2010.
- [22] Popp R L, Pattipati K R, “Bar-Shalom Y. m-Best S-D assignment algorithm with application to multitarget tracking”, *IEEE Trans. on AC*, Vol.37,No.1,pp.22 – 38, Jan,2001.

Authors' Profiles



Deqiang Zhou was born in 1975. He is a associate professor in the School of Information and Mathematics of Yangtze University of China interested in machine learning and wavelet analysis. He received his M.S.degree from Hubei University of China in 2004. His research interest include machine learning theory, grey systems theory and complicated mathematical models. His articles are published by Power System Protection and Control, Applied Mathematics, a Journal of Chinese Universities, Power system Technology, and others.