A Grey Model-Least Squares Support Vector Machine Method for Time Series Prediction

Ai WANG, Xuedong GAO

Abstract: In this study, the authors aim to solve the time series prediction problem through pre-predicting multiple influence factors of the target sequence. Focusing on two pre-prediction approaches of influence factors (i.e., factors driven approach and time driven approach), we propose a time series prediction method based on the least squares support vector machine and grey model (GM-LSSVM). This method could improve the prediction precision of the target time series by differentiating the data characteristics of influence factors. A case study is put forward to predict China's economy from the perspective of system innovation and technological innovation. We selected public statistics data from 2005 to 2014 from the national bureau. The numerical experiment results illustrate that the accuracy of the GM-LSSVM is able to reach 95%, which proves the effectiveness of our proposed method in practice.

Keywords: economic growth; grey model; least squares support vector machine; time series prediction

1 INTRODUCTION

Increasing scenarios of artificial intelligence technologies are keeping studies on time series prediction at center stage [8, 34]. Especially in the economic context, large amounts of economic time series data are produced every second, and various business scenarios like the economic growth prediction, economic policy making, and social development condition analysis, motivate the continuous improvement of prediction methods [35].

Referring to relevant research, Tab. 1 describes traditional prediction methods, including least squares support vector machine, multiple regression model, grey prediction model, BP neural network and combination models of the methods above [9-12]. It can be seen that compared to the multiple regression model and grey prediction model, both the least squares support vector machine (LSSVM) and BP neural network have advantages in predicting nonlinear data sequences. Although the multiple regression model could also apply mathematical expressions to alternate the correlation relationship with uncertainty between dependent and independent variables, further assumptions are needed for predicting nonlinear data sequences. Besides, the grey prediction model (GM) is able to predict the target data sequence with uncertainty through analyzing the correlation relationship between the development trends of system factors (influence factors of the target sequence).

Hence, this paper studies the time series prediction problem through pre-predicting the influence factors (of the target sequence). The main contributions are as follows. Firstly, this paper defines the pre-prediction approaches of influence factors (i.e, the factors driven approach and time driven approach). Secondly, according to the factors driven approach, a prediction method based on the LSSVM and GM (GM-LSSVM) is proposed. Thirdly, a case study on predicting China's economic growth is put forward using the statistics data among 2005 to 2014 from the national bureau. The influence factors of economic growth are also established through the grey correlation analysis theory. The numerical experiment results show that our proposed method GM-LSSVM has high precision in predicting the economic growth in practice.

Table 1 Comparison of traditional prediction methods

Prediction Method	Description	Characteristics
Least squares support vector machine (LSSVM)	Use nonlinear mapping algorithm to transform the linearly inseparable samples in low-dimensional input space into linear separable ones in a high-dimensional feature space, in order to construct classifiers.	Fit nonlinear samples well, have strong generalization ability, and do not rely on the distribution type of samples.
Multiple regression model	Use mathematical expressions to alternate the correlation relationship with uncertainty between dependent and independent variables.	Assumptions are needed for nonlinear problems.
Grey prediction model (GM)	Through analyzing the correlation relationship between the development trend of system factors, generate data sequences with strong regularity, and then establish and solve the differential equation model to complete the whole prediction process.	Be able to predict systems with uncertainty, but have more errors when doing medium or long-term prediction on data sequences which include abnormal points.
BP neural network	Through learning and storing a large number of input - output model mapping, use back propagation to constantly adjust the network weights and threshold, in order to minimize the sum of square error.	Fit nonlinear samples well with strong fault tolerance, but have local minimization problems and low convergence speed.

The remainder of the paper is organized as follows. In Section 2, after presenting the previous work related to our research (including the least squares support vector machine and grey model), we propose the prediction method GM-LSSVM. In Section 3, we conduct our experiment via the public dataset from 2005 to 2014 on the National Bureau of Statistics Data. The case study on predicting China's economic growth is also discussed. Finally, the paper is concluded in Section 4.

2 TIME SERIES PREDICTION BASED ON THE LSSVM AND GM

2.1 Pre-prediction Approaches for Time Series Prediction

According to the relationship between influence factors and the target sequence, the pre-prediction of target sequence could be divided into two different approaches (see Fig. 1): (1) Time driven approach (see Fig. 1a). The time driven approach utilizes the pre-predicted value of the target sequence at every historical time period to predict the target value. (2) Factors driven approach (see Fig. 1b). The factors driven approach utilizes the pre-predicted value of all the influence factors at the target time period to predict the target value.

In Fig. 1a, Y is the target object, X_i ($i = 1,2, …, n$) is the influence factor of Y, and the shadow areas on the left side represent the known data under the historical time period $(1,2, …, t)$, while the white areas on the right side represent the unknown data under the target time period $t + 1$, thus y_{t+1} is the target prediction value. The time driven approach will firstly pre-predict the target sequence at every historical time period $(y_1, y_2, y_3, …, y_t)$ using the influence factors $(x_{i1}, x_{i2}, x_{i3}, …, x_{it},)$ $(i = 1, 2, …, n)$, and then predict the target value y_{t+1} using the pre-prediction results $(y_1, y_2, y_3, …, y_t)$. If the real data of target object Y at every historical time period is available, that is $(y_1, y_2, y_3, …, y_t)$ is known data, the time driven approach could directly obtain the final result y_{t+1} through omitting the pre-predicting process.

Similarly, in Fig. 1b, The factors driven approach will firstly pre-predict all the influence factors at the target time period $(x_{1(t+1)}, x_{2(t+1)}, x_{3(t+1)}, …, x_{n(t+1)})$ using the influence factors $(x_{i1}, x_{i2}, x_{i3}, …, x_{it},)$ $(i = 1, 2, …, n)$, and then predict the target value y_{t+1} using the pre-prediction results$(x_{1(t+1)}, x_{2(t+1)}, …, x_{n(t+1)})$.
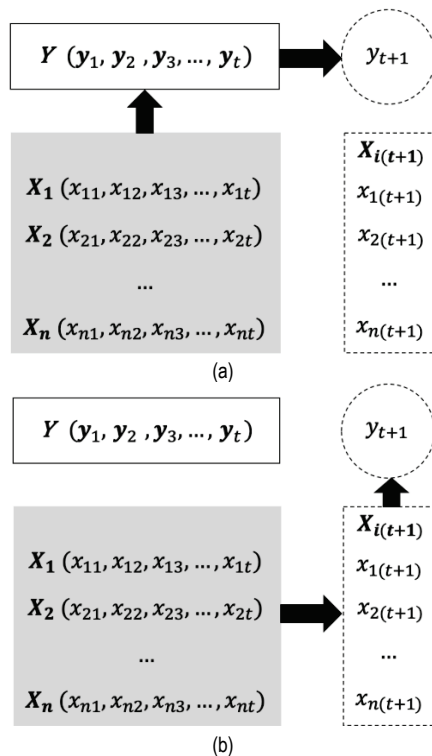

Figure 1 (a) The time driven approach of pre-prediction, (b) The factors driven approach of pre-prediction

2.2 Time Series Prediction Method Bases on the GM-LSSVM

Since there is much uncertainty in the collection of time series data among practical business scenarios, especially the economic data [14, 15], the least squares support vector machine (LSSVM) and grey prediction model (GM) are applied for the time series prediction (see Tab. 1).

The support vector machine model can well fit the nonlinear relationship between economic growth and its influence factors, through combining the structural risk minimization principle with prior risk and incredible risk,. Compared with other prediction methods, LSSVM has incomparable advantages on solving the finite sample, high dimension, local minimum and nonlinear problems [16, 17].

Moreover, the system which is consisted of economic growth and its influence factors is a grey system with great uncertainty, so it is difficult and unnecessary to find all the elements effecting economic growth. GM series model is the basic model of grey prediction theory, and takes the uncertainty system with little data as the research object. Especially the average GM (1, 1) model proposed by Deng. J has a wide range of application [18, 19].

Thus, focusing on the factors driven approach (see Section 2.1), this paper proposes the time series prediction method based on least squares support vector machine and grey prediction model (GM-LSSVM). The steps are as follows:

Input: 1) The time series of influence factors $X_i(t)$, ($i = 1, 2, …, n$; $t = 1, 2, …, M$); 2) The time series of target object $Y(t)$, ($t = 1, 2, …, M$); 3) The division proportion of training set and testing set μ.

Output: The prediction value of the target object $\hat{Y}(t + \Delta t)$, ($t = 1, 2, …, M$; $\Delta t \in N$).

Step 1: Calculate the prediction value $\hat{X}_i (t + \Delta t)$ if the influence factor $X_i(t)$ after Δt period of time prediction method.

i. Set the original sequence of influence factors as $X_i^{(0)}(t) = \left\{ x_i^{(0)}(1), x_i^{(0)}(2), ..., x_i^{(0)}(M) \right\}$, and calculate the new sequence $X_i^{(1)}(t) = \left\{ x_i^{(1)}(1), x_i^{(1)}(2), ..., x_i^{(1)}(M) \right\}$ through cumulating original sequence via Eq. (1).

$$x_i^{(1)}(k) = \sum_{t=1}^{k} x_i^{(0)}(t) \tag{1}$$

Where $k = 1, 2, …, M$.

ii. Establish the differential equation $\dfrac{\mathrm{d}X_i^{(1)}}{\mathrm{d}t} + aX_i^{(1)} = u$ on the cumulated sequence, and calculate the parameter a, u via the least square method, following Eq. (2):

$$\left[a_i, u_i \right]^{\mathrm{T}} = \left(B_i^{\mathrm{T}} B_i \right)^{-1} B_i^{\mathrm{T}} P_i \tag{2}$$

where,

$$B_i = \begin{bmatrix} -\dfrac{1}{2}\left[x_i^{(1)}(1)+x_i^{(1)}(2)\right] & 1 \\ -\dfrac{1}{2}\left[x_i^{(1)}(2)+x_i^{(1)}(3)\right] & 1 \\ -\dfrac{1}{2}\left[x_i^{(1)}(3)+x_i^{(1)}(4)\right] & 1 \\ \dots & \dots \\ -\dfrac{1}{2}\left[x_i^{(1)}(M-1)+x_i^{(1)}(M)\right] & 1 \end{bmatrix}, \ P_i = \begin{bmatrix} x_i^{(0)}(2) \\ x_i^{(0)}(3) \\ x_i^{(0)}(4) \\ \dots \\ x_i^{(0)}(M) \end{bmatrix}.$$

iii. Get the prediction model through calculating the differential equation above via parameter a, u following Eq. (3):

$$x_i^{(1)}(t+1) = \left(x_i^{(0)}(1) - \frac{u}{a}\right)e^{-at} + \frac{u}{a} \tag{3}$$

Get the reduction model Eq. (4) via Eq. (3) by derivation:

$$\hat{x}_i^{(0)}(t+1) = (-a)\left(x_i^{(0)}(1) - \frac{u}{a}\right)e^{-at} \tag{4}$$

Step 2: Train the prediction model.
i. According to the original data set $\{X_i(t), Y(t)\}$ ($t = 1, 2, \dots, M$) and division proportion, decide the train sample $S = \{(x_t, y_t) \mid (t = 1, 2, \dots, M)\}$, where M is the sample size, $x_t \in R^n$, $y_t \in R$, $x_t = \{X_i \mid i = 1, 2, \dots, \mu\}$, and $y_t = Y(t)$.

ii. Construct the regression function following Eq. (5):

$$y(x) = \sum_{t=1}^{M} \lambda_t K(x_j, x_t) + \theta \tag{5}$$

where,

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & \varphi(x_j)^{\mathrm{T}}\varphi(x_t)+\varepsilon^{-1} & \dots & \varphi(x_j)^{\mathrm{T}}\varphi(x_t) \\ \dots & \dots & \dots & \dots \\ 1 & \varphi(x_j)^{\mathrm{T}}\varphi(x_t) & \dots & \varphi(x_j)^{\mathrm{T}}\varphi(x_t)+\varepsilon^{-1} \end{bmatrix} \begin{bmatrix} \theta \\ \lambda_1 \\ \dots \\ \lambda_M \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \dots \\ y_M \end{bmatrix}.$$

Moreover, the purpose of the transformation above is to map the input space R^n to the high-dimensional feature space H via the nonlinear mapping function $\varphi'(*)$, and construct the optimal decision function among H. In (5), ε is the adjustment coefficient, θ is the deviation value and $\in R$, λ_r is the Lagrange operators and $\lambda_r \in R$.

iii. Select the kernel function. The kernel function and kernel parameter determine the prediction precision. The commonly used kernel functions are RBF kernel function, polynomial kernel function, and linear kernel function etc.
a) RBF kernel function

$$K(x_j, x_t) = \exp\left(\frac{-\|x_j - x_t\|^2}{\sigma^2}\right) \tag{6}$$

where, σ^2 is the width of the kernel function.

b) Polynomial kernel function

$$K(x_j, x_t) = \left(\tau + x_t^{\mathrm{T}} x_j\right)^d \tag{7}$$

c) Linear kernel function

$$K(x_j, x_t) = x_t^{\mathrm{T}} x_j \tag{8}$$

This paper selects the RBF function as the kernel function with strong generalization ability.
Step 3: Predict the target value after Δt period of time via the trained model and the prediction results of influence factors.
From the steps of GM-LSSVM method, it can be seen that the index system of influence factors has a great importance for the accuracy of prediction results. The ideal index system should be closely related to the target object, rather than find all the influence factors. For example, in the economic context, it is impossible to find all the influence factors of economic growth. Thus, this research will establish the influence factors of economic growth based on grey correlation analysis theory.
Julong Deng published the grey control system which marks the start of grey system theory [20]. Grey correlation analysis model establishes the correlation degree between elements through the similarity of curve geometry. According to the quantitative analysis, grey correlation degree can be calculated after comparing the geometrical relationship among time series. The comparative data series with the larger grey correlation degree is more similar to the reference data series on the developing direction and speed, and has the closer relationship with the reference sequence. The sample size of grey correlation analysis medal can be less than four. What's more, it can also be used for irregular data without the inconsistence between qualitative analysis results and quantitative results [21, 22].
The steps of grey correlation analysis algorithm are as follows:
Input: Time series; correlation coefficient $\rho (\rho = 0.5)$.
Output: Correlation degree.
Step 1: Set the analysis series. The reference series and comparative series should be set artificially. Reference series reflects the characteristic of the system behavior. Comparative series affects the behavior of the system. For example, regard $\{x_1^{(0)}\}$ as a mother sequence, and the rest of the series are subsequences.
Step 2: Initialize the original data. Because various system data may have different dimensions, it is difficult to get the correct conclusion when comparing with each other. So the grey correlation analysis algorithm generally must carry on the dimensionless processing of original data via Eq. (9).

$$x_i(k) = \frac{X_i(k)}{X_i(l)} \tag{9}$$

where, $k = 1, 2, \dots, n$; $i = 0, 1, 2, \dots, m$.
Step 3: Calculate the absolute difference between each subsequence and mother sequence via Eq. (10).

$$\Delta_{1i} = \left| x_1(t) - x_i(i) \right| \qquad (10)$$

Step 4: Calculate correlation coefficient via Eq. (11).

$$r_{1i} = \frac{\Delta_{\min} + \rho \Delta_{\max}}{\Delta_i(t) + \rho \Delta_{\max}} \qquad (11)$$

where, $i = 0, 1, 2, …, m$.

Step 5: Calculate correlation degree. Because the correlation coefficient is the correlation value between comparative series and reference series in each moment, every point in the curve, so its number is more than one. The information is too scattered to compare. Thus, it is necessary to concentrate the correlation coefficient in every moment, each point in the curve, to a value through calculating its average. The average is named as correlation degree between comparative series and reference series. The correlation formula is as follows:

$$R_{ij} = \sum_{t=1}^{n} r_{ij}(t) / n \qquad (12)$$

where, $t = 1, 2, …, n; i = 1, 2, …, s; j = 1, 2, …, t$.

Step 6: analyze correlation degree. According to the order of correlation degree, the larger correlation degree is, the greater the index affects the mother series. Last, get the final results through analyzing the mean and variance of the correlation degree.

3 EXPERIMENT ANALYSIS AND DISCUSSION
3.1 Experiment Design and Data Preparation

In this section, a case study on the economic growth prediction is put forward to verify the accuracy and effectiveness of the proposed method GM-LSSVM.

Economic growth usually refers to the continuous increase of output per capita (or income per capita) among a time span, which is the focus of all countries [1, 2, 33]. Through qualitative analysis, it can be found that economic growth is influenced by system innovation, technological innovation and industrial structure adjustment. The correlation of them is shown in Fig. 2 [3]. System innovation provides the development environment for

technological innovation, while technological innovation leads to the changes of system and rules. What's more, the results of system innovation and technological innovation will contribute to the adjustment of industrial structure. Because the industrial structure is stable in the short term, its effect on both others is relatively small [4-7, 31, 32]. So this paper mainly focuses on the influence of system and technological innovation on the economic growth.

Since economic growth is mainly affected by system and technological innovation, appropriate indicators should be selected to measure the economic growth, system innovation and technological innovation. According to the data released on the national bureau of statistics and related research, the following indexes in Tab. 2 are chosen as our initial index system [13, 23, 24].

This paper uses GDP to describe economic growth. GDP cannot only measure the overall national output and income scale, but also economic fluctuations and economic cycle state on the whole [25].

Number of published papers (X_{11}), published works (X_{12}) and authorized patents (X_{13}) reflect the national technological innovation level directly, while the high-tech products exports(X_{14}) and technical market turnover(X_{15}) reflect the developing progress of the technology innovation indirectly.

Taxes (X_{21}), Money supply (X_{22}), CPI (X_{23}), Foreign exchange reserve (X_{24}) and RMB exchange rate (X_{25}) are the indexes which are completely controlled by the state system; conversely the results of system innovation will ultimately reflect on these five indexes. So we chose the five indexes above as indicators describing system innovation.
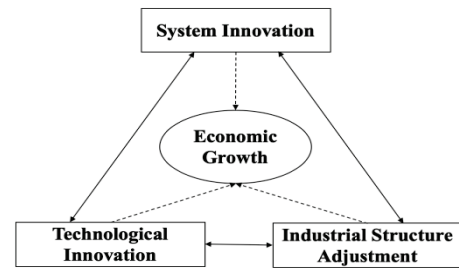

Figure 2 Three aspects that influence economic growth

Table 2 The influence factors of China's economic growth

Influence Factor	Description	Category of Influence Factors
Y	GDP (10^8 RMB)	Describing economic growth(Y)
X_{11}	Number of published papers	Describing technological innovation (X_1)
X_{12}	Number of published works	
X_{13}	Number of authorized patents	
X_{14}	The high-tech products exports (10^8 RMB)	
X_{15}	Technical market turnover (10^8 RMB)	
X_{21}	Taxes (10^8 RMB)	Describing system innovation(X_2)
X_{22}	Money supply (10^8 RMB)	
X_{23}	CPI	
X_{24}	Foreign exchange reserve (10^8 RMB)	
X_{25}	RMB exchange rate	

Tab. 2 contains 1 target object and 10 influence factors. There are two problems in the indicators selection process: (1) if we directly conduct the prediction with this 11 indicators, it will lead to large amounts of calculation, which affects the efficiency of LSSVM method [26]; (2) the selection of indicators is subjective, which can ensure the selected indexes and all the influence factors of

economic growth to belong to a sufficient relationship rather than necessary. Thus, it is significant to keep the chosen influence factors having a close relationship to economic growth with less number via reasonable streamlining the index system. The grey correlation analysis method in section 2.3 is able to accomplish this demand through seeking relevance to replace integrity.

According to the national bureau of statistics data, we got the original data of the 11 indexes above around 10 years (2005 - 2014) shown in Tab. 3.

It can be seen that in Tab. 3 there are great differences among the range of 11 indexes, so we should conduct the data preprocessing job first based on grey correlation analysis method and the initialization result is shown in Fig. 3.

Fig. 3 shows that each index overall has a rising trend after preprocessing, but the rising rate and amplitude are quite different.

After preprocessing, taking Y (GDP) as the mother sequence and other influence factors as subsequences, calculate the correlation degree between GDP and each influence factor in Tab. 4 via the results of absolute difference and correlation coefficient.

Table 3 The original dataset from the National Bureau of Statistics Data

Index	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Y	185895.8	217656.6	268019.4	316751.7	345629.2	408903	484123.5	534123	588018.8	636138.7
X_{11}	94.34	106.03	114.26	119.32	136.1	141.6	150	151.78	154.46	157
X_{12}	40120	42918	43063	45296	49080	45563	45472	46751	45730	47470
X_{13}	53305	57786	67948	93706	128489	135110	172113	217105	207688	233228
X_{14}	21824.8	28145	34780	41560.6	37693	49237.9	54880	60117	66030	66050
X_{15}	1551.37	1818.18	2226.53	2665.23	3039	3906.58	4763.56	6437.07	7469.13	8577.18
X_{21}	28778.5	34804.3	45621.9	54223.7	59521.5	73210.7	89738.3	100614.2	110530.7	119175.3
X_{22}	107278.8	126028.1	152560.08	166217.13	221445.8	266621.5	289847.7	308664.2	337291.05	348056.41
X_{23}	101.8	101.5	104.8	105.9	99.3	103.3	105.4	102.6	102.6	102
X_{24}	81887.2	106634	152824.9	194603	239915.2	284733.8	318114.8	331158.9	382131.5	384301.8
X_{25}	819.17	797.18	760.4	694.51	683.1	676.95	645.88	631.25	619.32	614.28

Table 4 The correlation degree result

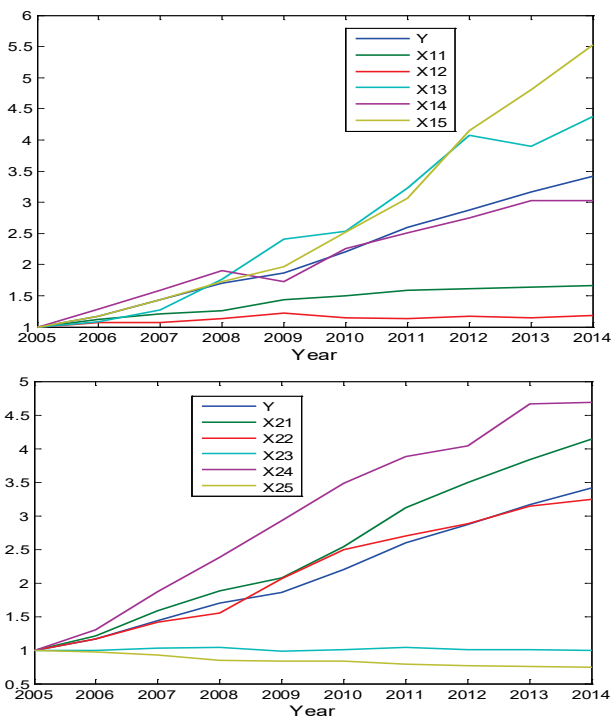| Index	R_{X11}	R_{X12}	R_{X13}	R_{X14}	R_{X15}	R_{X21}	R_{X22}	R_{X23}	R_{X24}	R_{X25}
Correlation Degree	0.70	0.63	0.77	0.91	0.78	0.81	0.94	0.61	0.64	0.58

Figure 3 The preprocessing results

Fig. 4 shows the sorting results of correlation degree in Tab. 4. It can be found that from the left to the right there are two obvious drops among the correlation curves on the whole, which can divide influence factors into four levels: Level 1 consists of X_{22}, X_{14} with $E(R_{x22}, R_{x14}) = 0.9$; Level 2 consists of X_{21}, X_{15}, X_{13} with $E(R_{x21}, R_{x15}, R_{x13}) = 0.8$; Level 3 consists of X_{11} with $E(R_{x11}) = 0.7$; Level 4 consists of X_{24}, X_{12}, X_{23}, X_{25} with $E(R_{x24}, R_{x12}, R_{x23}, R_{x25}) = 0.6$. Since the average correlation degrees of Level 3 and 4 are too small, remove them from the index system. Moreover, the correlation degrees of X_{15}, X_{13} are quite similar in Level 2, only keep X_{15} with larger degree. Above all, the final index system filtered by grey correlation analysis method only contains Money supply (X_{22}), the high-tech products exports (X_{14}), Taxes (X_{21}), and technical market turnover (X_{15}).

Tab. 5 shows the mean and variance of contribution degrees between system and technological innovation via Tab. 2. It can be seen that the mean of two-group contribution degree is similar, which means the influence of system innovation or technological innovation to economic growth is close. However, the variance of system innovation is much larger than the technological innovation. Thus, the results show that the contribution level of technological innovation to the economic growth is relatively stable, while reasonable system innovation will bring a surge in economic growth.

Table 5 The mean and variance of contribution degree

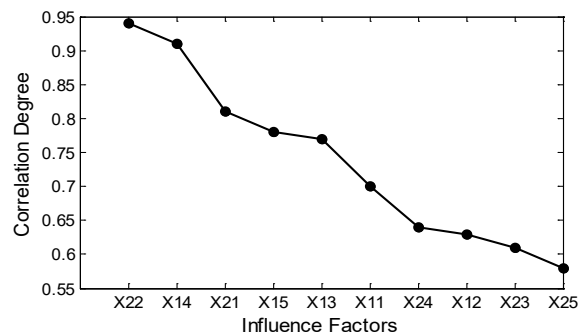| Index Category	Mean	Variance
Technological Innovation	0.75875	0.00856
System Innovation	0.71516	0.04752

Figure 4 The sorting result of correlation degrees

3.2 Experiment Results and Discussion

After establishing influence factors in section 3.1, Fig. 5 describes their original data from 2005 to 2014. It can be seen from Fig. 5 that a few time series are in "V" shape these 10 years, which reflects China's economic had

obvious fluctuations from 2008 to 2010. This section is to accomplish the prediction of China's economic growth based on the GM-LSSVM model in section 2.2.

Firstly, determine the input variables and parameters: 1) input time series of 4 influence factors $\{X_{22}(t), X_{14}(t), X_{21}(t), X_{15}(t) \mid t = 2005, 2006, …, 2014\}$. 2) input time series of the target object $Y(t)$, ($t = 2005, 2006, …, 2014$); 3) determine the division proportion $\mu = 0.7$, that is to say the train set contains data from 2005 to 2011, while the test set contains data from 2012 to 2014.

According to step 1, calculate the value of 4 influence factors based on the GM (1, 1) prediction method via the original data in previous seven years. For verification purpose，this experiment only completes the prediction

for the next three years. Tab. 6 shows the value of influence factors calculated by the GM (1, 1), and uses the relative error ((prediction value − actual value)/actual value) to measure the accuracy of prediction results. We can figure out that all the average relative error is beyond 0.05, and the average relative errors of system innovation indexes X_{22} and X_{21} are obviously higher than technological innovation indexes X_{14} and X_{15}. This is mainly because the uncertainty of innovation itself is big, and the subjectivity of system innovation is relatively stronger than technological innovation, which further verifies that only using the time series prediction model to predict data with strong uncertainty makes it hard to get ideal forecast results.

Table 6 Error Analysis of GM (1, 1) Prediction Model

Index	Actual Value			Prediction Value			Relative Error			E (Relative Error）
	2012	2013	2014	2012	2013	2014	2012	2013	2014	
X_{14}	6.0	6.6	6.6	6.1	6.8	7.5	0.02	0.03	0.14	0.06
X_{21}	1.0	1.1	1.2	1.1	1.2	1.3	0.05	0.09	0.13	0.09
X_{22}	3.1	3.4	3.5	3.5	3.8	4.1	0.15	0.12	0.17	0.15
X_{15}	6.4	7.5	8.6	5.7	7.5	9.2	-0.12	0.01	0.07	0.07

According to step 2, train the LSSVM model by the original data of target object and its influence factors in previous seven years, and test the trained model by the data in last 3 years. The results are shown in Fig. 6, where Y_{exp} represents the actual value of GDP, while Y_{pre} represents the prediction value of GDP by LSSVM trained model. The relative error of trained LSSVM model is 0.014, which has reached good performance.
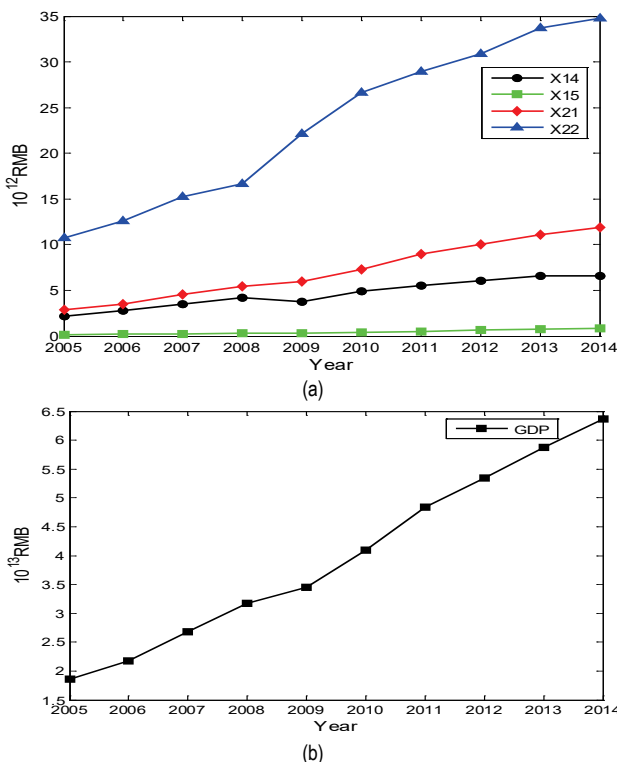

(a)

(b)

Figure 5 (a) The original data of influence factors, (b) The original data of target object

According to step 3, predict GDP in the last 3 years via trained LSSVM model and the prediction results of influence factors based on GM (1, 1). The results are shown in Tab. 7.

In order to test the accuracy of prediction results, this paper applies these error measurements which are widely used: relative error δ $\left(\delta = \dfrac{1}{N} \sum_{i=1}^{N} \dfrac{\left| y_i^{exp} - y_i^{pre} \right|}{y_i^{exp}} \right)$, absolute error MAE $\left(MAE = \dfrac{1}{N} \sum_{i=1}^{N} \left| y_i^{exp} - y_i^{pre} \right| \right)$ and the generalized root mean square error $RSME$ $\left(RSME = \sqrt{\dfrac{1}{N} \sum_{i=1}^{N} \left(y_i^{exp} - y_i^{pre} \right)^2} \right)$ [27], where y_i^{exp} represents the actual GDP value in the ith year, y_i^{pre} represents the prediction GDP value in the ith year, N is the sample size ($N = 3$, $i = 2012, 2013, 2014$).

According to the error analysis in Tab. 8, the performance of GM-LSSVM on three error measurement is better than GM (1, 1), and the improvement of accuracy ((error of GM (1, 1)-error of GM-LSSVM)/error of GM (1, 1)) is all over 50%. What's more, combining the results in Tab. 6, the relative error of the final target object predicted by GM-LSSVM is 0.05, which is smaller than all the influence factors.

The high precision of the GM-LSSVM model is mainly caused by the following reasons:

(1) GM (1, 1) adopts the ERM principle (empirical risk minimization) when doing parameters estimation, which leads to the low performance on the medium and long term prediction process with abnormal points. However, LSSVM is based on the SRM principle (structural risk minimization), which contributes to the feature of good nonlinear fitting, strong generalization ability, and independence of the sample distribution etc. So, GM-LSSVM combines two methods together by seeking compromise between complexity and generalization, and obtains better prediction results [28].

(2) Generally there are big leaps among all the original time series in Fig. 5. The other important difference between GM-LSSVM and GM (1, 1) is that GM-LSSVM uses multiple GM (1, 1) model to deal with the data

characteristics, and the essence of GM (1, 1) is to make the trend feature of time series more apparent through accumulating the original data, which improves the prediction efficiency [29, 30]. Thus, for the prediction problem, that target object time series and influence factors time series both have big leaps, separately dealing with the trend feature of individual time series and then reasoning out the whole value is better than directly dealing with the trend feature of the whole time series in the final prediction effect. (The definition of data leap characteristics and optimal prediction performance law of individuals with leaps will be presented in detail through another paper.)

Table 7 Prediction Results (105 million yuan)

Time	Actual Value	GM-LSSVM	GM (1, 1)
2012	5.3412	5.1769	5.5730
2013	5.8802	6.1362	6.4830
2014	6.3614	6.8176	7.5416

Table 8 Prediction Results (105 million yuan)

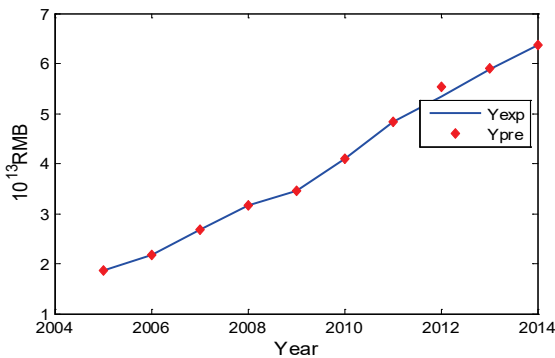| Error	GM-LSSVM	GM (1, 1)	Improvement of Accuracy
δ	0.05	0.11	54.5%
MAE	0.29	0.67	56.7%
RSME	0.32	0.78	59.0%

Figure 6 Trained results of LSSVM

4 CONCLUSIONS

According to the prediction problem of economic growth, this paper achieves the following results: (1) put forward the GM-LSSVM prediction model, which integrated the input-output prediction model and the time series prediction model together. GM-LSSVM overcomes the difficulty that LSSVM must know the value of influence factors first in order to predict the target object in a certain period of time. At the same time, the experiment results show that compared with GM (1, 1), the improvement of prediction accuracy on GM-LSSVM is more than 50%. (2) simplify the index system of influence factors through grey correlation analysis method, and gain the final prediction results of China's economic growth, whose precision has reached 95%.

The future research will mainly focus on increasing the accuracy of the GM-LSSVM prediction model. Although GM-LSSVM has achieved good prediction results on multiple error performance, there is still some room to improve. Moreover, more comparison experiments on the real economic data for different time periods will also be conducted using the GM-LSSVM to study the tendency of economic development.

5 REFERENCES

[1] Li, J. & Chen, X. (2015). Analysis of China's real estate prices and macroeconomy based on evolutionary co-spectral method. *Journal of Industrial Engineering and Management*, *8*(2), 684-688. https://doi.org/10.3926/jiem.1419

[2] Cheng, C. (2013). A study of dynamic econometric relationship between urbanization and service industries growth in china. *Journal of Industrial Engineering and Management*, *6*(1), 339-341. https://doi.org/10.3926/jiem.657

[3] Qi, W. (2010). Institutional change, technological innovation, structure adjustment and economic growth based on Russia. *Social Sciences Abroad*, *1*(1), 44-50.

[4] Liang, P. & Liu, H. (2007). The effects of indigenous innovation on economic growth in china: evidence from the provincial data. *The Proceedings of IE&EM'2007: Building Core Competencies through IE&EM*, *1*, 1238-1242.

[5] Wang, Q. & Cheng, L. (2009). The grey correlation analysis on enterprise technological innovation and economic performance. *Forum on Science and Technology in China*, *12*(1), 64-68.

[6] Yuan, Q. (2002). The comment on the correlation of technological and system innovation. *Academic Journal of Zhongzhou*, *127*(1), 51-53.

[7] Wang, A. (2005). The correlation analysis on technological, system and industrial innovation. *Contemporary Economic Research*, *08*(1), 31-34.

[8] Sun, Z. & Xu, L. (2015). The grey prediction of beijing economic growth. *Journal of Commercial Economics*, *10*(1), 131-133.

[9] Wang, S., Chen, A., Su, J., & Li, S. (2009). Application of the combination prediction model in forecasting the GDP of China. *Journal of Shandong University(Natural Science)*, *44*(2), 56-59.

[10] Li, D. & Li, Z. (2010). Research on predicting GDP in Guangxi province based on GM(1,1) and grey correlation analysis method. *Journal of Anhui Agricultural Sciences*, *38*(24).

[11] Teng, G. & He, Y. (2010). Research on China's GDP prediction using the combination method GMDH. *Statistics and Decision*, *7*(1), 17-19.

[12] Xiong, Z. (2011). Research on GDP time series forecasting based on integrating ARIMA with neural networks. *Journal of Applied Statistics and Management*, *30*(2), 306-314.

[13] Wei, N. (2010). Research of the time series analysis method and its application in forecasting the GDP of Shanxi province: Northwest A& F University Press.

[14] Sun, F. (2010). The method and application of grey correlation analysis. *Science & Technology Information*, *17*(1), 880-882.

[15] Wang, H., Hua, G., Jia, M., & Chen, Z. (2015). LSSVM and GM based dynamic soft sensor for coal level of ball mills. *Thermal Power Generation*, *44*(1), 77-81.

[16] Liao, X. (2016). Forecast model of road traffic accidents based on LS-SVM with grey correlation analysis. *Application Research of Computers*, *33*(3), 806-809.

[17] Geng, L., Zhang, T., & Zhao, P. (2012). Forecast of railway freigt volumes based on LSSVM with grey correlation analysis. *Journal of the China Railway Society*, *34*(3), 1-6.

[18] Wang, X., Zu, P., Zhao, B., & Zhao, W. (2013). Forecast of GDP in Mudanjiang province based on GM(1,1) and its influence factors analysis using grey correlation analysis method. *Mathematics in Practice and Theory*, *43*(8), 42-49.

[19] Yang, H., Liu, J., & Zheng, B. (2011). The improvement and application of grey prediction GM(1,1) model. *Mathematics in Practice and Theory*, *41*(23), 39-46.

[20] Liu, S., Xie, N., & Forrest, J. (2010). On new models of grey incidence analysis based on visual angle of similarity and nearness. *Systems Engineering*, *30*(5), 881-887.

[21] Tian, M., Liu, S., & Bu, Z. (2008). Research overview of the grey correlation analysis algorithm. *Statistics and Decision*, *253*(1), 23-27.

[22] Liu, W., Men, D., Liang, J., & Wang, W. (2012). Monthly load forecasting based on grey relational degree and least squares support vector machine. *Power System Technology*, *36*(8), 228-232.

[23] Li, W. & Wei, Y. (2012). Technology transfer, adaptation & assimilation and indigenous invention patent output: evidence from Chinese high-tech industries. *Procedia Engineering*, *29*(1), 1392-1398. https://doi.org/10.1016/j.proeng.2012.01.146

[24] Boko, C. E. F. (2013). Research on local company's innovation process based on knowledge transfer in developing countries: Wuhan University of Technology Press.

[25] He, X. (2012). Prediction of China's economic growth based on the time series. *The Theory and Practice of Finance and Economics*, *33*(178), 96-99.

[26] Tao, S., Chen, D., & Hu, W. (2007). Gradient algorithm for selecting hyper parameters of LSSVM in process modeling. *Journal of Chemical Industry and Engineering (China)*, *58*(6), 1514-1517.

[27] Ahmadi, M., Rozyn, J., Lee, M., & Bahadori, A. (2016). Estimation of the silica solubility in the superheated steam using LSSVM modeling approach. *Environment Progress & Sustainable Energy*, *35*(2), 596-602. https://doi.org/10.1002/ep.12251

[28] Zhou, D. (2016). Estimation of GM(1,1) model parameters based on LSSVM algorithm and application in load forecasting. *2011 International Conference on Engineering and Business Management*, *1*, 3538-3541.

[29] Wang, L. (2013). Models derived from GM(1,1) power model. *Systems Engineering-Theory & Practice*, *33*(11), 2894-2902.

[30] Liu, S., Zeng, B., Liu, J., & Xie, N. (2014). Several basic models of GM(1,1) and their applicable bound. *Systems Engineering and Electronics*, *36*(3), 501-508.

[31] Du, J., Li, Q., Qiao, F., & Yu, L. (2018). Estimation of vehicle emission on mainline freeway under isolated and integrated ramp metering strategies. *Environmental Engineering & Management Journal*, *17*(5), 1237-1248. https://doi.org/10.30638/eemj.2018.123

[32] Gunduz, M. & Khader, B. K. (2019). Management performance assessment: A method of analytic network process. *Journal of System and Management Sciences*, *9*(4), 67-90.

[33] Petrenko, Y., Masood, O., Javaria, K., & Vechkinzova, E. (2019). Entrepreneurship education and startups: A case study of less developed countries. *Journal of System and Management Sciences*, *9*(4), 39-49.

[34] Du, J., Li, Q., & Qiao, F. (2018). Impact of Different Ramp Metering Strategies on Vehicle Emissions Along Freeway Segments. *Journal of Transport & Health*, *9*, S51. https://doi.org/10.1016/j.jth.2018.05.036

[35] Wang, A. & Gao, X. (2019). Index System Reduction Method Based on the Index Similarity. *Technical Gazette*, *26*(4), 1112-1118. https://doi.org/10.17559/TV-20190615044223

Contact information:

Ai WANG, PhD, Student
University of Science and Technology Beijing,
No. 30 Xueyuan Road, Haidian District, Beijing, China
E-mail: wangai22222@126.com

Xuedong GAO, PhD, Professor,
(Corresponding author)
University of Science and Technology Beijing,
No. 30 Xueyuan Road, Haidian District, Beijing, China
E-mail: gaoxuedong@manage.ustb.edu.cn