ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

Chinese and American GDP Forecasts Based on Machine Learning

Hao Lv[1, a]

[1]Department of Statistics, School of Economic, Jinan University, Guangzhou, China.

[a]haolveric@163.com

Abstract

This article uses three machine learning methods, auto regressive integrated moving average model (ARIMA), BP neural network, support vector machine (SVM) to analyze the GDP between China and the United States from 1998 to 2017. Relative error is used to evaluate the accuracy of the proposed model, which can be used for GDP prediction. It is expected that this paper can provide some meaningful references for scholars who are studying in the field of GDP prediction. The machine learning methods, which include ARIMA time series, support vector machine SVM and BP neural network, are used to model and predict the GDP of the two countries. Comparison of the three methods finally concluded that the BP neural network is most suitable for the use of GDP prediction analysis.

Keywords

GDP forecast; Machine learning; ARIMA, BP, SVM.

1. INTRODUCTION

As we all know, GDP is an important indicator to measure the level of a country's economic development. Predicting one country's GDP in advance can provide important information for social resource allocation and national policy decisions. Predicting GDP with the technology of machine learning has gradually become a research focus. Shao Xiaorui[1] et al. explained how to combine three machine learning methods of ARIMA, BP and SVM to predict the relevant indicators in the paper of "Traffic Accident Time Series Prediction Model Based on Combination of ARIMA and BP and SVM". However, GDP is a different type of data. Therefore, in this paper, the data related to GDP are used for modeling and prediction. General economic prediction models include mathematical and statistical regression method [2-3], time series prediction method [4-5], Markov chain method [6], grey prediction model [7], support vector machine, neural network method as well as other nonlinear regression method. In this paper, BP neural network, SVM and ARIMA are used to construct the model and predict future GDP of China and the United States of America.

2. PROPERTIES

(1) Auto Regressive Integrated Moving Average Model, ARIMA

ARIMA Model (Auto Regressive Integrated Moving Average Model), also known as Integrated Moving Average auto-regressive Model, is one of the time series prediction analysis methods. In ARIMA (p, d, q), AR is "auto regression", and p is the number of auto regression terms. MA is the "moving average", q is the number of moving average terms, and d is the degree of difference to make it a stationary sequence. [9]

The ARIMA (p, d, q) model is an extension of the ARMA (p, q) model. ARIMA (p, d, q) model can be expressed as:

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$$
$$d \in \mathbb{Z}, d > 0.$$

L is the lag operator.

(2) Support Vector Machine SVM

Support Vector Machine (SVM) is a generalized linear classifier that classifies data in binary classification according to supervised learning, and its decision boundary is the maximum-margin hyper plane [10-12]. In this paper, regression is constructed, so the support vector machine regression is mainly introduced.

If SVM is extended from Classification problem to Regression problem, Support Vector Regression (SVR) can be obtained. At this time, the standard algorithm of SVM is also known as Support Vector Classification (SVC). The hyper plane decision boundary in SVC is the regression model of SVR:

$$f(\boldsymbol{X}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{X} + b$$

If the sample point is close enough to the regression model, that is, it falls into the interval boundary of the regression model, then the corresponding loss function is called insensitive loss function:

$$L(z) = \max(0, |z| - \epsilon)$$

ϵ is the super parameter that determines the width of the interval boundary. It can be seen that the insensitive loss function is similar to the hinge loss function used by SVC, and the value near the origin is fixed at 0. SVR is a quadratic convex optimization problem in the following form:

$$\begin{aligned} \max_{w,b} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 \\ \text{s.t.} \quad & \left|y_i - f(\boldsymbol{X})\right| \le \epsilon \end{aligned}$$

The variable ξ, ξ^* is used to represent the piecewise value of the insensitive loss function.

$$\begin{aligned} \max_{w,b} \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \\ \text{s.t.} \quad & y_i - f(\boldsymbol{X}) \le \epsilon + \xi_i \\ & f(\boldsymbol{X}) - y_i \le \epsilon + \xi_i^* \\ & \xi \ge 0, \quad \xi^* \ge 0 \end{aligned}$$

Similar to SVM, through the Lagrange multiplier:

$$\boldsymbol{\alpha}, \boldsymbol{\alpha}^*, \boldsymbol{\mu}, \boldsymbol{\mu}^*$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

The Lagrange function and dual problem can be obtained:

$$\begin{aligned}\mathcal{L}(w,b,\xi,\xi^*,\alpha,\alpha^*,\mu,\mu^*) &= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i+\xi_i^*\right) - \sum_{i=1}^{N}\mu_i\xi_i - \sum_{i=1}^{N}\mu_i^*\xi_i^* \\ &\quad + \sum_{i=1}^{N}\alpha_i\left[f(\boldsymbol{X}_i)i - y_i - \epsilon - \xi_i\right] + \sum_{i=1}^{N}\alpha_i^*\left[f(\boldsymbol{X}_i)i - y_i - \epsilon - \xi_i^*\right]\end{aligned}$$

$$\max_{\alpha,\alpha^*} \quad \sum_{i=1}^{N}\left[y_i\left(\alpha_i^*-\alpha_i\right) - \epsilon\left(\alpha_i^*+\alpha_i\right)\right] - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\left(\alpha_i^*-\alpha_i\right)(\boldsymbol{X}_i)^{\mathrm{T}}\left(\boldsymbol{X}_j\right)\left(\alpha_j^*-\alpha_j\right)\right]$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\left(\alpha_i^*-\alpha_i\right) = 0, \quad 0 \le \alpha_i, \alpha_i^* \le C$$

The dual problem has the following KKT condition:

$$\begin{cases} \alpha_i\alpha_i^* = 0, \quad \xi_i\xi_i^* = 0 \\ (C-\alpha_i)\xi_i = 0, \quad \left(C-\alpha_i^*\right)\xi_i^* = 0 \\ \alpha_i\left[f(\boldsymbol{X}) - y_i - \epsilon - \xi_i\right] = 0 \\ \alpha_i^*\left[y_i - f(\boldsymbol{X}) - \epsilon - \xi_i^*\right] = 0 \end{cases}$$

By solving the dual problem, the form of SVR can be obtained as follows:

$$f(\boldsymbol{X}) = \sum_{i=1}^{m}\left(\alpha_i^*-\alpha_i\right)\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{X} + b$$

SVR can obtain nonlinear regression results by kernel method.

(3) BP neural network

BP (Back Propagation) neural network is a concept proposed by Rumelhart and Mcclelland in 1986. It is a multi-layer forward neural network trained by the algorithm of error backward propagation, which is the most widely used neural network at present [13].

The artificial neural network does not need to determine the mathematical equation of the mapping relation between input and output in advance. Through its own training and learning rules, it can get the result closest to the expected output value when the input is given. As an intelligent information processing system, BP neural network is a multi-layer forward network trained by error back propagation. Its algorithm is called BP algorithm. Its basic idea is gradient descent method.

The basic BP algorithm includes forward propagation of signal and backward propagation of error. In other words, the error output is calculated in the direction from input to output, while the weight and threshold are adjusted in the direction from output to input. In the case of forward propagation, the input signal acts on the output node through the hidden layer and generates the output signal through nonlinear transformation. If the actual output is inconsistent with the expected output, the error is transferred into the process of backward propagation. Error backward propagation means that the output error is transmitted back to the input layer through the hidden layer, and the error is distributed to all the units of each layer.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

The error signal obtained from each layer is used as the basis for adjusting the weight of each unit. By adjusting the connection intensity between the input node and the hidden layer node, not only the connection intensity between the hidden layer node and the output node, but also the error drops along the gradient direction. After repeated learning and training, the network parameters, such as weight and threshold value are determined, and the training will stop. At this time, the trained neural network can process the information with minimal output error by itself for the input information of similar samples.

3. TESTS

In this paper, the differential integration moving average auto regressive model (ARIMA), BP neural network and SVM model of support vector machine are adopted. Different models have different requirements on data input. We will perform empirical study based on machine learning algorithm as follow.

(1) Data description

This paper adopts the GDP data of China and the United States from 1998 to 2017. The data is from China's national statistics website [14]. The data does not include missing values, so there is no need to fill or replace the missing values. In addition, the data unit scale is consistent with only one column for both countries, so there is no need to do the standardization process. The specific data are shown in table 1 below.

Table 1. Raw data (in trillion dollars)

Year	China	U.S.A.
1998	1.029	9.0892
1999	1.094	9.0006
2000	1.2113	10.2848
2001	1.3394	10.6218
2002	1.4706	10.9775
2003	1.6603	11.5106
2004	1.9553	12.2749
2005	2.286	13.0937
2006	2.7521	13.8559
2007	3.5222	14.4776
2008	4.5982	14.7186
2009	5.11	14.4187
2010	6.1006	14.9644
2011	7.5726	15.5179
2012	8.5605	16.1553
2013	9.6072	16.6915
2014	10.4824	17.4276
2015	11.0647	18.1207
2016	11.191	18.6245
2017	12.2377	19.3906

Visualize the data for both countries, the results are shown in figure1 below

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

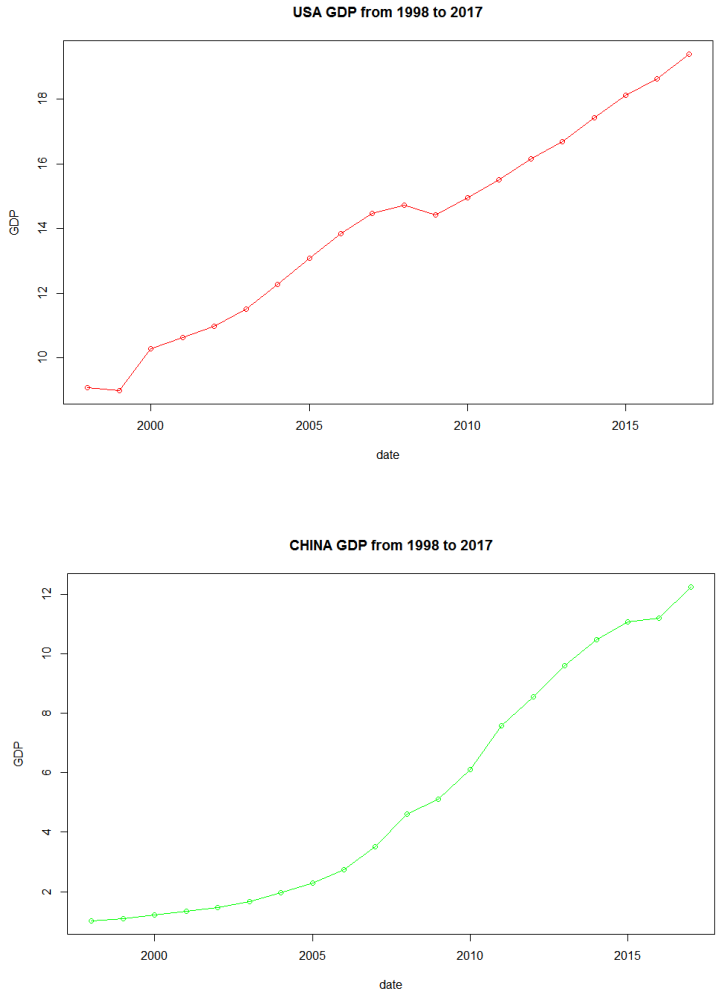

Figure 1. GDP of the United States and China from 1998 to 2017

It can be seen from the figure that GDP of both countries has a growth trend from 1998 to 2017. GDP growth of China is basically exponential, while that of the United States is relatively slow. Placing the two paths on one diagram as shown in figure 2, it is obvious that China was relatively small compared to the United States. Although China has grown rapidly since 2012, the gap between two countries is still huge.

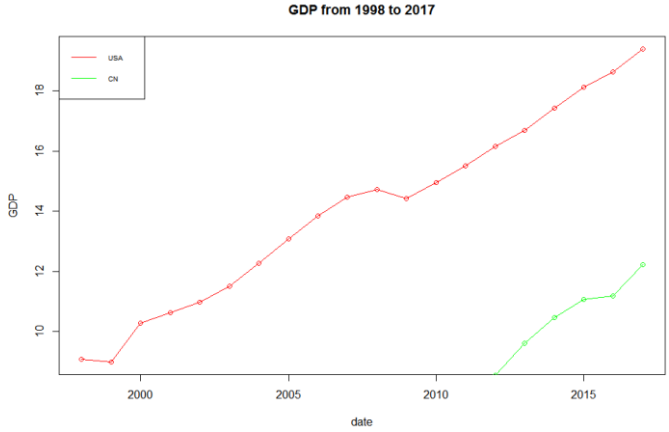

Figure 2. Comparison of U.S. and Chinese GDP from 1998 to 2017

(2) Model construction

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

Method 1: Construction of average auto regressive model (ARIMA)

Step 1 White noise test:

In order to build the ARIMA model, white noise test is firstly conducted on the data. The white noise test can determine whether the original data is meaningful for further research or not. Take China's GDP data as an example. The p-value obtained is far less than 0.05, indicating that the original data is not white noise, thus can be further analyzed and processed.

Step 2 Stationary test:

The requirement of ARIMA model is that the data must be stable. However, as shown in FIG. 1, both GDP data of China and the United States have an upward trend, so it is necessary to conduct differential processing. Through the experiment, the second order difference processing can obtain stationary time series. The results are shown in figure 3.

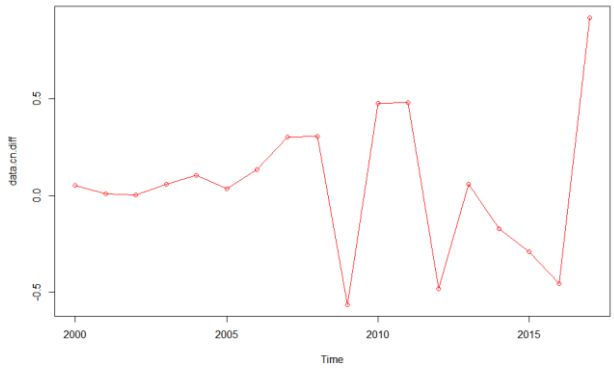

Figure 3. Stationary time series after second-order difference processing

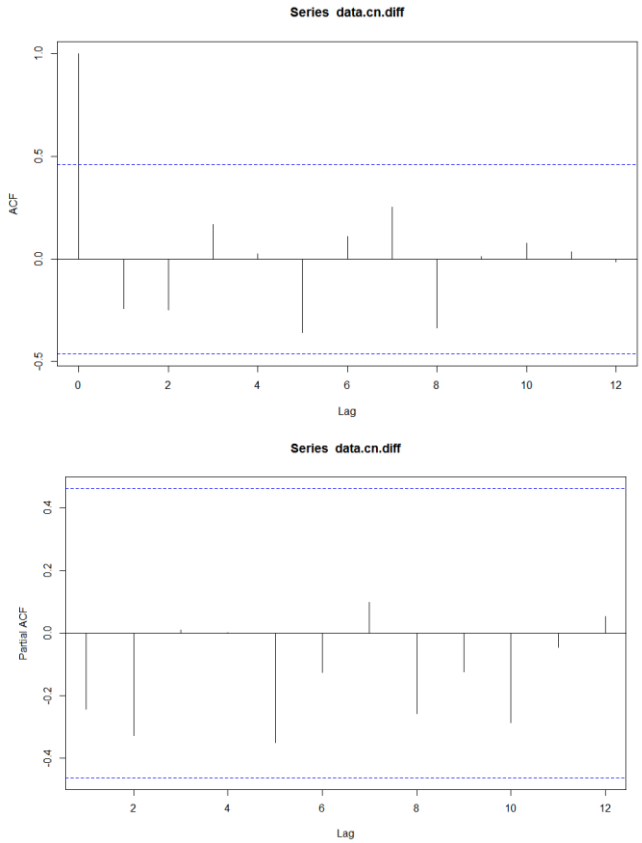

Figure 4. Test results of ACF and PACF

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

Step 3 Determination of acf and pacf:

In order to determine the values of p and q, acf and pacf are tested in the definite stage. Since acf shows that the data of order 0 fluctuates within twice the standard deviation, the value of p is determined to be 0. Similarly, pacf fluctuates within twice the standard deviation after order 0, so ARIMA model is determined as ARIMA (0, 2, 0), where 2 is the difference of second order. The results of acf and pacf tests are shown in figure 4.

Step 4 Fitting results:

The established ARIMA (0,2,0) model is used to fit the data, and the results are shown in figure 5. It can be seen from the figure that the data model fitted by ARIMA (0,2,0) is relatively good.

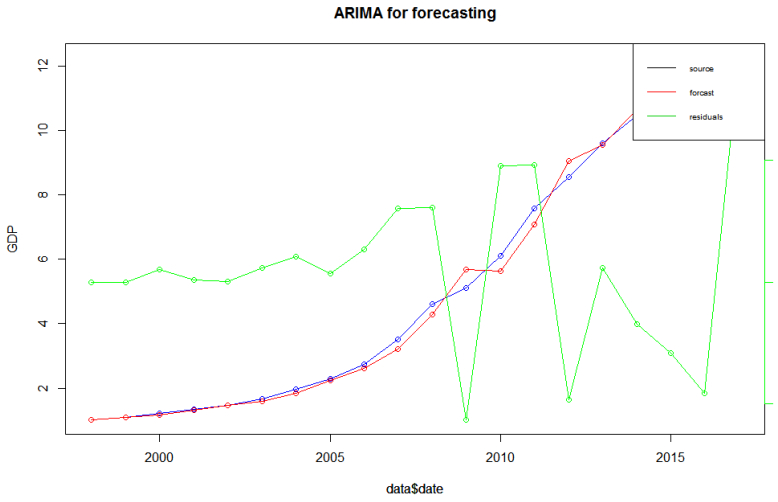

Figure 5. Comparison of ARIMA (0,2,0) fitting results with the original data and the obtained residuals

Method 2: BP neural network modeling

The BP neural network needs to determine the number of layers including input layer, hidden layer and output layer. This paper adopts the method of 0.618, and the number of layers is determined by the following method. The model adopts the method of input layer 5 and output layer 1.

$$m = \begin{cases} n + 0.618(n - t), n \geq t \\ n - 0.618(t - n), n < t \end{cases}$$

In the equation above, m represents the number of nodes in the hidden layer, and t represents the number of nodes in the output layer. And the definition of BP input node is shown in table 2.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

Table 2. BP neural network input and output

Year	Input $x = x[x'(t-5), x'(t-4), x'(t-3), x'(t-2), x'(t-1)]$					Output
Array	$x'(t-5)$	$x'(t-5)$	$x'(t-5)$	$x'(t-5)$	$x'(t-5)$	$x'(t)$
1998	0.0000	0.0000	0.0000	0.0000	0.0000	1.029
1999	0.0000	0.0000	0.0000	0.0000	1.029	1.094
2000	0.0000	0.0000	0.0000	1.029	1.094	1.2113
2001	0.0000	0.0000	1.029	1.094	1.2113	1.3394
2002	0.0000	1.029	1.094	1.2113	1.3394	1.4706
2003	1.029	1.094	1.2113	1.3394	1.4706	1.6603
2004	1.094	1.2113	1.3394	1.4706	1.6603	1.9553
...	...	...	...	...	...	...
2013	4.5982	5.11	6.1006	7.5726	8.5605	9.6072
2014	5.11	6.1006	7.5726	8.5605	9.6072	10.4824
2015	6.1006	7.5726	8.5605	9.6072	10.4824	11.0647
2016	7.5726	8.5605	9.6072	10.4824	11.0647	11.191
2017	8.5605	9.6072	10.4824	11.0647	11.191	12.2377

According to the table, we use the data of previous five years to predict the data of the current year. The prediction results are shown in figure 6. It can be seen from the figure that the fitting of BP algorithm is almost exactly the same as the original data, and its residual varies between [-0.04, 0.04].

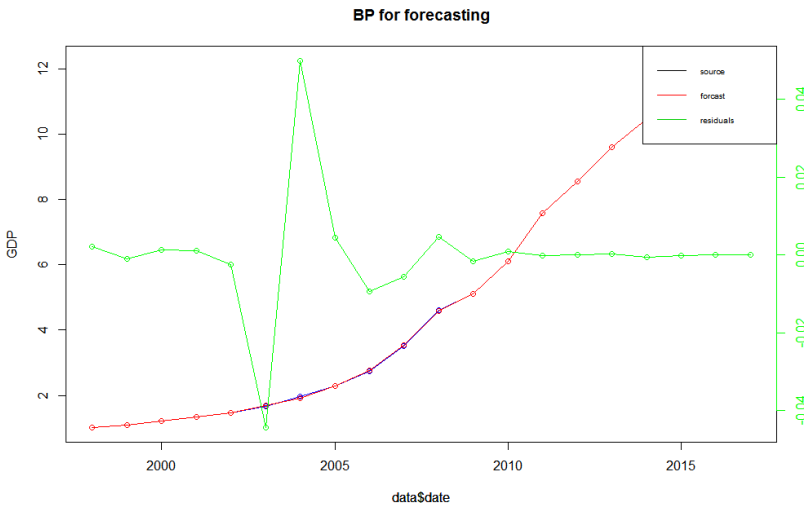

Figure 6. BP results

Method 3: The SVM model

SVM modeling is processed in the same way as BP algorithm, the results are shown in figure 7. The residual of results varies between [-0.10,0.10], thus the effect of SVM is not as good as BP.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

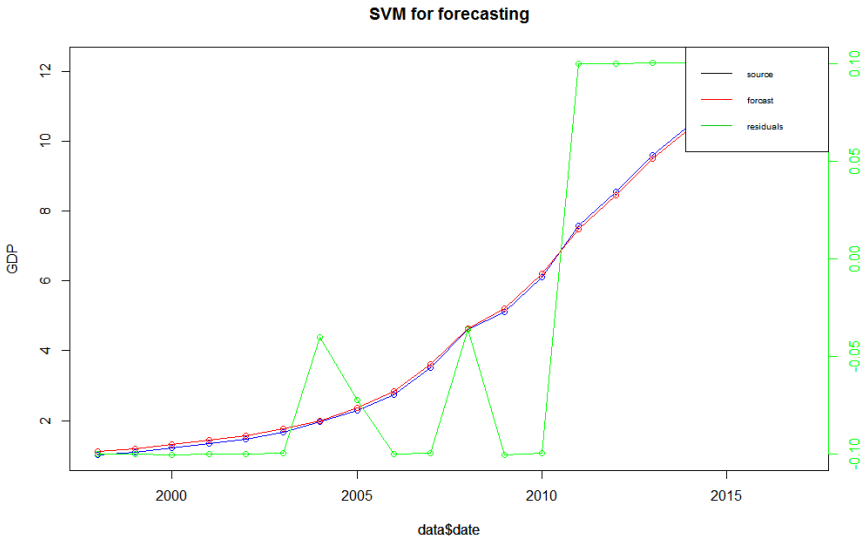

Figure 7. SVM results

4. COMPARISON AND CONCLUSION

From the above analysis, we can basically conclude that the fitting effect of BP algorithm is the best. Finally, we need to evaluate the model from the quantitative results. Table 3 lists the specific predictive fitting values of the three models. As results can be seen from the table, the performance of BP neural network is the best, followed by ARIMA. The performance of SVM is not as good as the other two methods.

Table 3. Comparison of ARIMA (0,2,0), BP and SVM

Year	China	U.S.	BP	SVM	ARIMA error	BP error	SVM error
1998	1.029	1.028539817	1.026900065	1.028539817	4.00E-04	0.002	-0.0973
1999	1.094	1.095235201	1.094917229	1.095235201	-0.0011	-8.00E-04	-0.0916
2000	1.2113	1.159	1.210013503	1.159	0.0432	0.0011	-0.0828
2001	1.3394	1.3286	1.338375051	1.3286	0.0081	8.00E-04	-0.0746
2002	1.4706	1.4675	1.472966358	1.4675	0.0021	-0.0016	-0.0681
2003	1.6603	1.6018	1.704610162	1.6018	0.0352	-0.0267	-0.0601
2004	1.9553	1.85	1.905392044	1.85	0.0539	0.0255	-0.0205
2005	2.286	2.2503	2.281471346	2.2503	0.0156	0.002	-0.0316
2006	2.7521	2.6167	2.761451422	2.6167	0.0492	-0.0034	-0.0363
2007	3.5222	3.2182	3.527892214	3.2182	0.0863	-0.0016	-0.0283
2008	4.5982	4.2923	4.593454191	4.2923	0.0665	0.001	-0.0078
2009	5.11	5.6742	5.111613674	5.6742	-0.1104	-3.00E-04	-0.0196
2010	6.1006	5.6218	6.099783903	5.6218	0.0785	1.00E-04	-0.0163
2011	7.5726	7.0912	7.572861677	7.0912	0.0636	0	0.0132
2012	8.5605	9.0446	8.56051791	9.0446	-0.0566	0	0.0117
2013	9.6072	9.5484	9.60686735	9.5484	0.0061	0	0.0104
2014	10.4824	10.6539	10.48289983	10.6539	-0.0164	0	0.0096
2015	11.0647	11.3576	11.06476627	11.3576	-0.0265	0	0.009
2016	11.191	11.647	11.19085783	11.647	-0.0407	0	0.0089
2017	12.2377	11.3173	12.23772052	11.3173	0.0752	0	0.0082

Through experiments and comparison of the three machine learning methods, it is found that different methods have different advantages in fitting and predicting the GDP data. Even though

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202006_6(6).0015

ARIMA method requires data stability as well as white noise test, it is the simplest among the three methods. BP and SVM algorithms need to reconstruct the original data, but the predicted results are relatively more accurate than ARIMA. The relative error of BP algorithm is the smallest among the three algorithms. Therefore, we consider BP neural network to be the best method compared to the other two. In conclusion, BP neural network is more suitable for GDP data prediction as could be seen by comparing the overall process of the three different machine learning methods based on the Chinese and American GDP data.

REFERENCES

[1] Shao Xiaorui, chang-soo Kim,Traffic Accident Time Series Prediction Model Based on Combination of ARIMA and BP and SVM, 2018 international conference of enviroment and computer science. 2018.9. Xiamen, China.

[2] GuyonI, Stork DG. Linear discriminant and support vector classifiers [M] SmolaAJ, BartlettPL ScholkopfB, et al. Advances in Large Margin Classifiers. Cambridge, MA:MIT Press, 2000.

[3] BurgeC. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121 167.

[4] Zhao Q, Principe Support vector machines for SAR automatic target recognition [J]. IEEE Transom Aerospace and Electronic Systems, 2001, 37 (2): 634 654.

[5] Kim K I, Jung K. Support vector machines for texture classification [J]. IEEE Transom Pattern Analysis and Machine Intelligence, 2002, 24 (11): 1542 1550.

[6] Ei-NaqaI, Yang YY. A support vector machine approach for detection of microcalcifications [J]. IEEE Transom Medical Image, 2002, 21 (12): 1552 1563.

[7] Fung GM, Mangasarian OL. Breast tumor susceptibility to chemotherapy via support vector machines, Technical Report 03 06 [R].Data Mining Institute, 2003.

[8] Huatu education. Administrative vocational ability test: red flag press, 2014.

[9] Percival, Donald B.; Walden, Andrew T. (1993). Spectral Analysis for Physical Applications. Cambridge University Press. ISBN 0-521-35532-X.

[10] Vapnik, V. Statistical learning theory. 1998 (Vol. 3). New York, NY: Wiley, 1998: Chapter 10-11, pp.401-492.

[11] Zhou Zhihua. Machine learning. Beijing: Tsinghua university press, 2016: pp.121-139, 298-300.

[12] Li Hang. Statistical learning methods. Beijing: Tsinghua university press, 2012: chapter 7, pp.95-135.

[13] Wen Xin, Zhang Xingwang, Zhu Yaping, Li Xin. Intelligent fault diagnosis technology:Matlab Application: Beijing university of aeronautics and astronautics press, 2015.09.