Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (22): 5132-5137, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.5132.5137

Forecasting Stock Volatility using LSSVR-based GARCH Model Optimized by Siwpso Algorithm

[1]Li-Yan Geng and [2]Fei Yu
[1]School of Economics and Management,
Shijiazhuang Tiedao University, Shijiazhuang, 050043, China
[2]CNPC Offshore Engineering Company Limited, Tanggu, Tianjin, 300451, China

Abstract: Volatility forecasting plays an important role in derivatives pricing, risk management and securities valuation. As a traditional parametric model, GARCH can't forecast financial volatility well. To improve the forecasting accuracy and the modeling speed of GARCH model, this study proposed a hybrid forecasting model for stock volatility forecasting, in which Least squares support vector regression (LSSVR), combining with stochastic inertia weight PSO (SIWPSO), is proposed to GARCH model. First, LSSVR was used to forecast financial volatility under the GARCH framework. Then, the SIWPSO algorithm was adopted to obtain the optimal hyper-parameters needed in the LSSVR model. An empirical research was performed to illustrate the effectiveness of the proposed method. Empirical results from four high frequency stock indices returns in China stock market indicate that the proposed model provides improvement in volatility forecasting performance. The values of HRMSE, HMAE, LL and LINEX of the proposed model are obviously smaller than those of the other two models: LSSVR-GARCH-CV and GARCH. The searching time for the optimal hyper-parameters of the LSSVR model by SIWPSO is much shorter than that under 10-fold CV method. Therefore, the proposed model is an effective approach for forecasting stock volatility.

Key words: Volatility forecasting, least squares support vector regression, stochastic inertia weight particle swarm optimization

INTRODUCTION

The forecasting of volatility in financial time series has been widely studied by scholars for many years. Since Engle (1982) and Bollerslev (1986) developed Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, it has been recognized that GARCH model and its various extensions are effective methods for volatility forecasting. To further improve the forecasting performance of the GARCH model, Artificial Neural Network (ANN) was applied to GARCH model for volatility forecasting (Schittenkopf *et al*., 2000; Tseng *et al*., 2008). Due to the flexibility of ANN to approximate the nonlinear relationship between past return and future volatility, ANN-based GARCH model outperforms the traditional parametric GARCH model in volatility forecasting. However, the traditional ANN also suffers from itself weakness such as getting trapping into multiple local optima, the curse of dimensionality and over-fitting.

Support Vector Machines (SVM), based on statistical learning theory (Vapnik, 1995), emerged as an effective method for financial time series forecasting. SVM implements the Structural Risk Minimization (SRM) principle which can theoretically obtain better forecasting performance than the traditional ANN. The applications of SVM to GARCH model for volatility forecasting have shown significant reduction in forecasting errors (Tang *et al*., 2009; Gavrishchaka and Banerjee, 2006; Chen *et al*., 2010).

Least Squares Support Vector Machines (LSSVM) is a modified version of SVM. Different from SVM, LSSVM gives the solution by a linear system instead of a quadratic programming problem (Suykens and Vandewalle, 1999) which decreases the complexity of calculation and raises the learning rate (Zhang, 2012; Zhang and Li, 2012). Least Squares Support Vector Regression (LSSVR) is another form of LSSVM for regression problem and has been successfully applied to the volatility forecasting (Ou and Wang, 2010; Geng *et al*., 2013). The selection of hyper-parameters is important in the forecasting performance of LSSVR. Cross validation is the common method for selecting the hyper-parameters of LSSVR. As a grid searching method, cross validation

Corresponding Author: Li-Yan Geng, School of Economics and Management, Shijiazhuang Tiedao University, Shijiazhuang, 050043, China

method could cause large amount workload and the hyper-parameters obtained may not be the best. That will have ill influence on forecasting accuracy. Stochastic Inertia Weight Particle Swarm Optimization (SIWPSO) algorithm is an improved PSO algorithm. In SIWPSO, the inertia weight is randomly selected based on a uniform distribution in a certain range which improves the search precision of the algorithm.

The objective of this study is to construct LSSVR based GARCH model with SIWPSO algorithm. Using data of China stock market, the performance of this method was evaluated by comparing it with two models: LSSVR based GARCH model with cross validation method and GARCH model.

MATERIALS AND METHODS

GARCH model: Let y_t be a time-series of bounded variance (e.g., returns on an index) and Φ_{t-1} be the information set of all information up to time t-1. Then a typical GARCH (1, 1) model for the time series y_t takes the form:

$$\begin{cases} y_t = \varepsilon_t h_t \quad \varepsilon_t \mid \Phi \sim \text{iid}(0,1) \\ h_t^2 = \kappa + \delta y_{t-1}^2 + \eta h_{t-1}^2 \\ t = 1, 2, \ldots, T \end{cases} \tag{1}$$

where, ε_t is a series of independent, identically distributed random variables with zero mean and unit variance. And h_t^2 is conditional variance. It can be seen from the definition in Eq. 1 that the conditional variance h_t^2 is a stochastic process expressed as a linear function of the last period of conditional variance, h^2_{t-1} and the last period of squared observation, y^2_{t-1}. Parameters: κ, δ and η are real parameters and the conditions: $\kappa>0$, $\delta\geq0$, $\eta\geq0$ are important to guarantee that the process h^2_t remains positive. The sum, $\delta+\eta$, is called the persistence parameter. If $\delta+\eta<1$, the process y_t is stationary.

Least squares support vector regression: LSSVR applies a squared loss function instead of the traditional quadratic programming method. In the primal weight space, the LSSVR is formulated as:

$$\text{Min} \quad \psi(\omega, e) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{i=1}^{n} e_i^2 \tag{2}$$

Subject to the equality constrains:

$$\bar{\theta}_i = \omega^T \varphi(\theta_i) + b + e_i \tag{3}$$

where, i = 1, 2,…,n, ω is weight vector and b is bias term; e_i is the error variable and γ is the regularization parameter; $\theta_i \in R^d$ and $\bar{\theta}_i \in R$ are the input vector and corresponding output variable, respectively. A set of nonlinear transformations $\varphi(\theta_i)$ allow LSSVR to work in a linear space.

The lagrangian function is constructed to solve the above optimization problem. According to Karush-Kuhn-Tucker condition, the following linear equations can be obtained as:

$$\begin{bmatrix} 0 & 1^T \\ 1 & \Omega + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{\theta} \end{bmatrix} \tag{4}$$

where, $a = [\alpha_1, \alpha_2, \ldots, \alpha_n]^T$, $? = [\bar{\theta}_1, \bar{\theta}_2, \ldots, \bar{\theta}_n]^T$, I is an identity matrix with n orders and $1 = [1,\ldots,]^T$. Kernel function matrix Ω has the elements with the form $\Omega_i = \varphi(\theta_i)^T \varphi(\theta) = K(\theta_i, \theta)$ for i = 1, 2,…, n. Finally, the LSSVR model becomes:

$$\bar{\theta}_i = \sum_{i=1}^{n} \alpha_i K(\theta_i, \theta) + b \tag{5}$$

where, α_i(i = 1, 2,…, n) are Lagrange multipliers and $K(\theta_i, \theta)$ is the kernel function satisfying the Mercer's condition.

LSSVR-based GARCH model: The traditional GARCH model is usually estimated by the maximum likelihood estimation (MLE) which requires specified assumptions distribution of the data beforehand. However, it is difficult to specific the function form and distribution of the data accurately.

Theoretically, LSSVR can flexibly approximate any continuous and nonlinear system well, without a priori distributional assumptions about the data. Motivated by this characteristic, LSSVR is applied to GARCH model and LSSVR-based GARCH (LSSVR-GARCH) model is expressed as follows:

$$h_t^2 = f\left(h_{t-1}^2, y_{t-1}^2\right) \tag{6}$$

where, h_t^2 is the output variables and $[h_{t-1}^2, y_{t-1}^2]$ is the input variables. Nonlinear function relationship f() is built by LSSVR through the training samples. From Eq. 5 and 6, the LSSVR-GARCH model for one-step-ahead volatility forecasting is obtained as:

$$h_{t+1}^2 = \sum_{t=1}^{n} \alpha_t K(h^2, h_t^2) + b \tag{7}$$

Both h_t^2 and $[h_{t-1}^2, y_{t-1}^2]$ must be known first in order to establish and train the LSSVR-GARCH model. Here, h_t^2 is reflected as:

$$h_t^2 = \frac{1}{5}\left(y_t^2 + y_{t-1}^2 + y_{t-2}^2 + y_{t-3}^2 + y_{t-4}^2\right) \tag{8}$$

The function which satisfies the Mercer's condition can be selected as kernel function in Eq. 7. In this study, RBF is considered because of the high nonlinear feature existed in the volatility series. The RBF kernel function has the following form:

$$K(h^2, h_t^2) = \exp\left(-\|h^2 - h_t^2\|^2 / 2\sigma^2\right) \tag{9}$$

where σ^2 is the kernel parameter. Consequently, according to Eq. 4 and 9, two hyper-parameters, (γ, σ^2), need to be selected appropriately in LSSVR model beforehand.

Parameters optimization of LSSVR by SIWPSO: The SIWPSO algorithm is applied to optimize the hyper-parameters (γ, σ^2) in order to improve the forecast accuracy of LSSVR-GARCH model. The inertia weight of the SIWPSO is a series of random numbers which follows a certain distribution. By changing the inertia weight, the global and local search capacities are dynamically adjusted. According to (Hu and Zeng, 2006), the formula of the stochastic inertia weight is:

$$\begin{cases} w = \mu + \lambda \cdot \mathrm{Var}(0,1) \\ \mu = \mu_{min} + \zeta \cdot (\mu_{max} - \mu_{min}) \end{cases} \tag{10}$$

where, μ and λ are the mean and variance of the inertial weight, μ_{max} and μ_{min} are the maximum and minimum average value of the inertial weight, respectively. And Var(0,1) is the random number followed by normal distribution in the range [0,1]. And Var(0,1) is the random number in the range [0,1].

The detailed procedures of optimizing the hyper-parameters (γ, σ^2) of LSSVR-GARCH by SIWPSO algorithm (LSSVR-GARCH-SIWPSO) for forecasting volatility are given below.

Step 1: Data preprocessing. The whole dataset are initially normalized using the mean and standard deviation of each variable

Step 2: Particle swarm initialization. Generate initial M sets of particles which consist of the hyper-parameters (γ, σ^2). Set the parameters including the maximum and minimum average value of the inertia weight, the variance of the inertia weight, the number of maximal iterations and acceleration coefficients

Step 3: Fitness function definition. The fitness function is forecasting error of the model, defined as:

$$\mathrm{Fitness} = \frac{1}{l}\sum_{t=1}^{l}(y_t^2 - \hat{h}_t^2)^2 \tag{11}$$

where, y_t^2 represents the actual volatility and $\hat{h}_t^2$ represents the forecasted volatility. And l represents the number of training samples

Step 4: Particles evolutionary. Calculate each particle's fitness values according to (11) and take the particle with the minimal fitness values in the swarm as the best position. The inertia weight *w* is updated according to (10)

Step 5: Stop criterion judgment. If the number of maximal iterations is reached, the evolutionary process is terminated. By this, the algorithm gives the optimal hyper-parameters (γ^*, σ^{2*}), otherwise, k = k+1, go back to step 3

Step 6: LSSVR establishment and forecasting. LSSVR model is established with the obtained optimal hyper-parameters and is used to forecasting volatility. Then, the forecasted volatility is transformed into the original volatility forecasts

RESULTS AND DISCUSSION

Data description: The data examined in the empirical research consist of one-minute high frequency close price on December 3, 2007 including Shanghai Composite Index (SHCI), Shanghai Stock Exchange 180 Index (SH180), HuShen300 Index (HS300) and A Share Index (ASI). These high frequency stock indices p_t are transformed into high frequency returns y_t by the following expression:

$$y_t = 100 \times (\log p_t - \log p_{t-1}) \tag{12}$$

There are 240 data samples for each return series. Each whole data sample is divided into two subsets: initial 170 samples are used to establish and train the model and the remaining 70 samples for forecasting volatility. Table 1 lists the descriptive statistics for four stock indices returns.

Table 1: Descriptive statistics of the one-minute stock index returns

Statistics	SHCI	SH180	HS300	ASI
Mean	0.0030	0.0056	0.0052	0.0029
Median	0.0024	-0.0008	0.0032	0.0011
Max	0.2162	0.1869	0.1778	0.2166
Min	-0.3703	-0.2384	-0.2628	-0.3729
S.D.	0.0647	0.0590	0.0558	0.0648
Skewness	-0.3463	0.2209	0.0605	-0.3649
Kurtosis	8.1238	4.2342	5.1505	8.2359
J-B	267.3300*	17.1828*	46.3922*	279.4748*
$Q^2(12)$	66.0277*	97.9864*	82.4480*	63.463*

*Denotes significantly at the 5% level. 2. Max, Min denote Maximum and Minimum, respectively; SD denotes Standard deviation. 5. J-B is the Jarque-Bera normality test; $Q^2(12)$ is the Ljung-Box Q test for 12th order serial correlation of the squared returns

Table 2: HRMSE, HMAE, LL, LINEX and TIME of out-of-sample volatility forecasts by three models

	Indices	HRMSE	HMAE	LL	LINEX($\times 10^{-6}$)	TIME(s)
SHCI	LSSVR-GARCH-SIWPSO	31.9370	10.7485	3.5745	6.0162	6.1719
	LSSVR-GARCH-CV	36.5113	12.0064	3.7492	6.3361	7182.9063
	GARCH	34.2603	13.6702	4.2293	6.0645	20.8750
SH180	LSSVR-GARCH-SIWPS-O	268.4733	67.7681	6.5027	18.4096	6.3906
	LSSVR-GARCH-CV	268.6599	65.8251	6.3476	18.5860	7846.7188
	GARCH	571.4139	100.4464	6.8528	19.2080	19.8906
HS300	LSSVR-GARCH-SIWPSO	300.6154	46.1326	4.1398	8.0564	6.5156
	LSSVR-GARCH-CV	313.7218	48.6043	4.2626	8.8149	8726.4688
	GARCH	840.7540	109.0940	4.4379	7.2242	19.3594
ASI	LSSVR-GARCH-SIWPSO	25.1163	9.1282	3.3800	5.9327	6.8281
	LSSVR-GARCH-CV	27.6935	9.4712	3.4471	6.3424	7866.9219
	GARCH	27.8209	11.5327	3.9469	6.1234	21.7500

The positive return means with nearly zero standard deviation are presented in the four return series. The distributions of SHCI and ASI are negatively biased (Skewness<0) and the distribution of SH180 and HS300 are positively biased (Skewness>0). The kurtosis values of the four return series are all bigger than 3 which suggest all return series have fat-tailed distributions. For four return series, the J-B normality tests reject the original normality hypothesis. The statistics $Q^2(12)$ show significant linear serial correlation in the squared returns.

Empirical results: In LSSVR-GARCH-SIWPSO model, the parameters of the SIWPSO algorithm are set as follows: $M = 10$, $\mu_{max} = 0.9$, $\mu_{min} = 0.1$, $c_1 = c_2 = 2$, λ and $k_{max} = 30$. The hyper-parameters of LSSVR are simulated continuously by the SIWPSO algorithm for ten times and the obtained optimal hyper-parameters are used to establish the GARCH-LSSVR model. Then the well trained model is used to one-step-ahead forecast the volatility of the four stock returns.

The volatility is also forecasted using LSSVR-GARCH-CV model and GARCH model. The out-of-sample volatility forecasting results of these two models are compared with those of the LSSVR-GARCH-SIWPSO model. In LSSVR-GARCH-CV model, the LSSVR is also used to estimate the parameters of GARCH model, but the hyper-parameters in LSSVR are selected by the 10-fold cross validation (10-fold CV) method. In GARCH model, the parameters are estimated by MLE, in which the stock indices returns are assumed to follow the normal distribution.

The forecasting performance of the three models is evaluated by four statistical criteria. These are the heteroscedasticity adjusted root mean squared error (HRMSE), the mean absolute error (HMAE), the logarithmic error statistic (LL) and the linear-exponential (LINEX). These statistical criteria are expressed as follows:

$$\text{HRMSE} = \left[N^{-1} \sum_{t=1}^{N} (1 - \hat{h}_t^2 / y_t^2)^2 \right]^{1/2} \tag{13}$$

$$\text{HMAE} = N^{-1} \sum_{t=1}^{N} \left| 1 - \hat{h}_t^2 / y_t^2 \right| \tag{14}$$

$$\text{LL} = N^{-1} \sum_{t=1}^{N} \left[\ln(\hat{h}_t^2) - \ln(y_t^2) \right]^2 \tag{15}$$

$$\text{LINEX} = N^{-1} \sum_{t=1}^{N} \{ \exp\left[a(\hat{h}_t^2 - y_t^2) \right] - a(\hat{h}_t^2 - y_t^2) - 1 \} \tag{16}$$

where, $\hat{h}_t^2$ is the volatility forecasts of different models and y_t^2 is the realized volatility measured by the squared returns. And N is the number of the volatility forecasts. The forecasting performance is better when the above criterion value is smaller. In addition, the searching time for the optimal hyper-parameters (TIME) is recorded to evaluate the convergence speed of the three models. The results are presented in Table 2.

Results discussion: According to Table 2, with the exclusion of LINEX of HS300, in all the four indices LSSVR-GARCH-SIWPSO model has the smallest values of HRMSE, HMAE and LL among the three models. As for LINEX of HS300, the second-smallest value is founded in the LSSVR-GARCH-SIWPSO model and the smallest value is in the GARCH model. This shows that as a whole, LSSVR-GARCH-SIWPSO model performs better than the two other models on volatility forecasting. It is clear from TIME that the searching time for the optimal hyper-parameters of LSSVR using SIWPSO is obviously shorter than that of the 10-fold CV method. This indicates that SIWPSO is superior to the cross validation method in terms of parameter estimation efficiency.

Ou and Wang (2010) combined LSSVM with GARCH-type models as a hybrid method to forecast stock volatility and found the improved performances of the hybrid method. But the proposed method is a

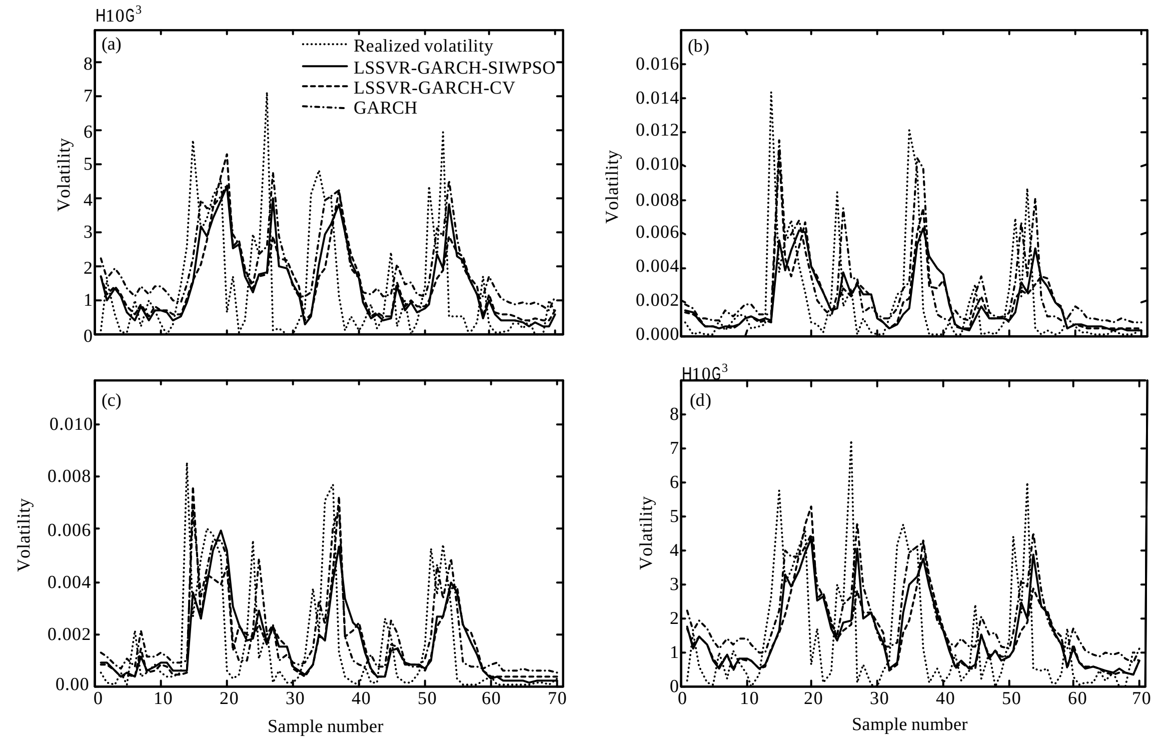

Fig. 1(a-d): Comparison of three stock index volatility forecasts for different models. (a) Comparison of SHCI stock index volatility forecasts for three models, (b) Comparison of SH180 stock index volatility forecasts for three models, (c) Comparison of HS300 stock index volatility forecasts for three models and (d) Comparison of ASI stock index volatility forecasts for three models

semi-parametric method and needs to estimate more parameters than LSSVR-GARCH-SIWPSO model which increases the computation complexity.

Figure 1a-d give the graphs of the one-step-ahead volatility forecasts from the three models for four stock indices. These figures show that LSSVR-GARCH-SIWPSO model forecasts the up change tendency of the realized volatility well and captures the general up peak in realized volatility better.

CONCLUSION

This study has forecasted the volatility of GARCH model based on LSSVR with SIWPSO algorithm. As demonstrated in the empirical research, the proposed model forecasts the volatility of four high frequency stock indices more accurately than the LSSVR-GARCH-CV model and the GARCH model. Furthermore, the convergence speed of the SIWPSO to LSSVR-GARCH model is much faster than that of the 10-fold CV method. In this study, the RBF function has been selected as the kernel function of LSSVR model and the SIWPSO has been used to optimize the parameters of LSSVR-GARCH model. To further improve the forecasting accuracy of the LSSVR-GARCH model, further works can focus on selecting other nonlinear kernels as the kernel function of LSSVR and using other improved intelligent optimization algorithms to adjust the optimal parameters of LSSVR-GARCH model.

ACKNOWLEDGMENT

This study was supported by the Scientific Research Foundation of the Ministry of Education of China for Young Scholars (11YJC790048).

REFERENCES

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econ., 31: 307-327.

Chen, S., W.K. Hardle and K. Jeong, 2010. Forecasting volatility with support vector machine-based GARCH model. J. Forecast., 29: 406-433.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50: 987-1007.

Gavrishchaka, V.V. and S. Banerjee, 2006. Support vector machine as an efficient framework for stock market volatility forecasting. Comput. Manage. Sci., 3: 147-160.

Geng, L.Y., F. Yu and X.P. Zhou, 2013. Grey least squares support vector machines with particle swarm optimization for volatility forecasting. Adv. Inform. Sci. Serv. Sci., 5: 580-588.

Hu, J.X. and J.C. Zeng, 2006. A particle swarm optimization model with stochastic inertia weight. Comput. Simul., 23: 164-167.

Ou, P. and H. Wang, 2010. Financial volatility forecasting by least square support vector machine based on GARCH, EGARCH and GJR models: Evidence from ASEAN stock markets. Int. J. Econ. Finance, 2: 51-64.

Schittenkopf, C., G. Dorffner and E.J. Dockner, 2000. Forecasting time-dependent conditional densities: A semi non-parametric neural network approach. J. Forecast., 19: 355-374.

Suykens, J.A.K. and J. Vandewalle, 1999. Least squares support vector machine classifiers. Neural Process. Lett., 9: 293-300.

Tang, L.B., H.Y. Sheng and L.X. Tang, 2009. GARCH prediction using spline wavelet support vector machine. Neural Comput. Appl., 18: 913-917.

Tseng, C.H., S.T. Cheng, Y.H. Wang and J.T. Peng, 2008. Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices. Phys. A Stat. Mech. Appl., 387: 3192-3200.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York, USA.

Zhang, M. and G. Li, 2012. Network intrusion detection based on least squares support vector machine and chaos particle swarm optimization algorithm. J. Convergence Inform. Technol., 7: 169-174.

Zhang, T.W., 2012. Study on the network attack prediction algorithm based on chaos theory and LSSVM. J. Convergence Inform. Technol., 7: 224-231.