International Journal of u- and e- Service, Science and Technology
Vol.9, No. 4 (2016), pp.141-152
http://dx.doi.org/10.14257/ijunesst.2016.9.4.14

Forecasting Algorithm Parameters Solving Method based on Phase-space Reconstruction and Its Application in Price Prediction

Hu Shuyu[1] and Huang Ming [2, *]

[1] *College of Network Education, Hunan Radio and Television University, Changsha, 410004, China*
[2] *College of Economics and Management, Hunan Radio and Television University, Changsha, 410004, China*

Abstract

To improve the forecasting accuracy of oil prices, this paper has proposed oil price predicting model (PSR-LSSVM) based on unified solving by phase space reconstruction and predicting algorithm parameters using interrelation between phase-space reconstruction and predicting algorithm. The LSSVM is selected as the predicting algorithm of oil prices, and the parameters of phase space reconstruction and LSSVM are taken as individuals of the genetic algorithm, and the optimal delay time, embedding dimension and LSSVM parameters are obtained through selection, crossover and mutation evolutionary mechanism, and finally, the predicting model of oil prices is established and the performance of predicting model is tested by Daqing oil price time series. The results show that the proposed model PSR-LSSVM obtains higher predicting accuracy than the oil-price forecasting models of independently optimized phase-space reconstruction and LSSVM, which provides a new research idea for the predicting problem of chaotic time series.

Keywords: *Oil prices; LSSVM; Phase-space reconstruction; Unified solution; Modeling prediction*

1. Introduction

As a major energy source, oil plays a vital role in a country's political stability and economic development and price of oil is affected by factors such as military, politics, economy and diplomacy synthetically,characterized by randomness, mutability, chaos and other characteristics of change. Therefore,oil price forecast has been a hot topic in the field of prediction [1].

Oil price forecasts based on chaos theory mainly include phase-space reconstruction and prediction algorithms, and both are interrelated and mutually influenced, used together in the forecasted results of oil price forecast model. In order to improve the accuracy of oil price forecasts, making overall consideration of the intrinsic link between the phase space reconstruction and prediction algorithm, this article has proposed oil price forecast model (PSR-LSSVM) based on unified solving by phase space reconstruction and predicting algorithm parameters, and has conducted simulation experiment of model performance using Daqing oil price data to test the feasibility and superiority of the proposed model.

[*] Corresponding Author

ISSN: 2005-4246 IJUNESST

2. Related Research

Oil prices have been proven to be a low-dimensional chaotic time series [2]. Phase-space reconstruction is basis of prediction for chaotic time and can tap evolution law hidden in chaotic attractors. In phase space reconstruction, selection of time delay (τ) and embedding dimension (m) is critical, directly affecting the prediction accuracy of the subsequent oil prices [3]. For the selection of τ and m parameters, there are mainly two ideas: ① τ and m is solved respectively. If τ is firstly determined using autocorrelation method and mutual information method, *etc*., then m is determined using G-P method, pseudo- nearest neighbor method. However, autocorrelation method can only extract linear correlation between time sequences. Although mutual information method can reflect multiple integrality relations of the system, these relationships are just projection of high-dimensional phase space in two-dimensional space, only reflecting the independence of the two adjacent reconstructed coordinates in reconstructed phase space, while they cannot guarantee the overall independence between all reconstructed coordinates. And in most of the methods, in order to determine the optimal τ, we need to determine m first of all, while τ shall be fixed in advance to determine m, which will inevitably lead to conflicts [4]. ② τ and m is uniformly solved. For instance, Broomhead *et al*. have put forward time window method; Kim *et al*. have proposed CC law that τ and m is uniformly solved using the relationship between τ and m, so that the reconstructed oil price time series can more accurately reflect the trend of oil prices, thus the idea of the unified solution often used for τ and m [5]. Currently, NN, SVMs and other machine algorithms [6-7] are mainly employed in the oil price prediction algorithms. Since LSSVM has advantages such as training speed and excellent generalization ability, *etc*., this paper has selected LSSVM as oil price prediction algorithm, while the forecast performance of LSSVM is closely related the selection of its parameters [8]. In oil price forecasting model based on LSSVM, its prediction accuracy is jointly determined by the phase space reconstruction and LSSVM, but the current oil price forecasts do not take account of the link between the two, completely separating the intrinsic link between the two, resulting in mismatching of τ and m selection with LSSVM parameter, thus it is difficult to establish the oil price forecasting model with optimal overall performance.

3. Phase-Space Reconstruction and LSSVM

3.1. Phase Space Reconstruction Theory

Assuming that time sequence is {xi}, its time interval of ∆t (unit time), and then for the power system with n variables, there is:

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \cdots, x_n) \tag{1}$$

By elimination method, it is converted into an n- order nonlinear differential equation:

$$x^{(n)} = f\left(x, x^{(1)}, \mathbf{n}, x^{(n-1)}\right) \tag{2}$$

After the conversion, the new track is:

$$X(t) = \left(x(t), x^{(1)}(t), \cdots, x^{(n-1)}(t)\right) \tag{3}$$

Formula (3) describes the dynamic system homeomorphous with the original time series, and it constitutes phase space reorganization by x (t) plus (n-1) order derivative $x(1)(t)$, $x(2)(t)$,…,$x(n\text{-}1)(t)$ to replace this continuous variable x (t) and its derivatives, without considering discontinuous time series and its time-lag displacement of $X(t) = (x(t), x(t+\tau), \cdots, x(t+(n-1)\tau))$ when n-1.

As long as the time delay τ is selected as the time scale of time series, linearly independent delayed coordinates can be ensured.

Assuming univariate time series is$\{x(ti), i=1,2,\ldots,n\}$, time delay of $\tau = k\Delta t, k = 1, 2, \cdots$, and then the time series can be extended to m -dimensional phase space:

$$X_i(t) = (x(t_i), x(t_i+\tau), \cdots, x(t_i+(m-1)\tau)) \tag{4}$$

In the formula, Xi (t) is a phase point of m -dimensional phase space.

For any phase point Xi (t), there are m components, and m satisfies the condition:

$$m = n - (m-1)k \tag{5}$$

A phase point in m -dimensional phase space represents state of the system after an instant, while the ligature of phase points constitute locus of points in phase space, and this "trajectory" would mean that the system state evolves over time. Then we get a phase type in m -dimensional phase space, and in principle, there would be enough information to reveal the dynamic characteristics of the time series in the multi-dimensional phase space to inspect [9].

According Takens theorem, the " trajectory" in the embedded space of phase-space reconstructed by suitable embedding dimension m and time delay τ is equivalent in dynamics to the original system, and accordingly:

$$X(t+T) = f(X(t)) \tag{6}$$

In the formula,, T is forecasting step and f () is the reconstructed prediction model.

According to equation (6), f () satisfying the formula (6) can be obtained through the known time series to get a predictive model.

3.2. LSSVM

For the training set $\{(xi,yi)\}, i=1,2,\ldots,n, xi$ xi and *yi* denote sample input and output respectively, and the samples are mapped into high-dimensional feature space via nonlinear mapping function Φ (), thereby obtaining the optimal linear regression function:

$$f(x) = w^T \phi(x) + b \tag{7}$$

In the formula, ω is the weight vector of feature space, and b is the offset value.

According to the structural risk minimization principle, the LSSVM regression model for problem solving in formula (7) is:

$$\begin{aligned} &\min \|w\|^2 + \frac{1}{2}\gamma \sum_{i=1}^{n} \xi_i^2 \\ &s.t. \\ &y_i - w^T \phi(x) + b = e_i \end{aligned} \tag{8}$$

In the formula, γ is regularization parameter, while ei is error.

By introducing the Lagrange multiplier, equation (8) is converted into the dual space optimization problem, namely:

$$\begin{aligned} L(w,b,\xi,\alpha) = {} & \min \|w\|^2 + \frac{1}{2}\gamma \sum_{i=1}^{n} \xi_i^2 + \\ & \sum_{i=1}^{n} \alpha_i (w^T \phi(x) - b + e_i - y_i) \end{aligned} \tag{9}$$

In the formula, α_i is the Lagrange multiplier.

According to Mercer condition, nuclear function is defined as follows:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{10}$$

RBF kernel function is selected as LSSVM kernel function and RBF kernel function is defined as follows:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{11}$$

In the formula, σ is the width of the kernel function[10].

Final LSSVM regression model is:

$$f(x) = \sum_{i=1}^{N} \alpha_i \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) + b \tag{12}$$

In modeling process of LSSVM, the parameters γ and σ produce impact on predicted forecasted results. Current LSSVM algorithm fails to consider relation between γ and σ as well as τ and m, thus there are some defects in the established prediction model. To overcome this drawback, this paper has proposed thinking of uniform solution to τ and m and γ and σ.

4. Unified Solving by Parameter of Oil price Forecast Model

Unified solution of T, m and γ, σ is a multi-parameter combinatorial optimization problem, and genetic algorithms (GA) as bionic algorithm can find the global optimal solution in a short time [11] due to its advantages such as intelligent search, parallel computing and global optimization and others, therefore this paper performs unified solution of τ, m and γ, σ using GA algorithm.

4.1. Individual Coding Design

To facilitate crossover and mutation operations, GA individual uses binary encoding, and each individual includes four parts of τ, m and γ, σ. Because individual employs binary coding, in conducting forecasts and modeling of oil prices, parameters are converted to decimal numeral through formula (13).

$$p = \min_p + \frac{\max_p - \min_p}{2^l - 1} \times d \tag{13}$$

In the Formula, p represents the decimal value of the parameter; minp and maxp represent the minimum and maximum parameter; l represents individual binary length; d represents the decimal value of binary string of the parameter.

4.2. Fitness Function Design

The quality of the particles is evaluated by the fitness function value, and the aim of unified solving by phase space reconstruction and prediction algorithm parameters is to improve the prediction accuracy in oil prices, thus oil price forecasting accuracy is employed as the fitness function of particles, that is:

$$f(x) = accuracy(\tau, m, \gamma, \theta) \tag{14}$$

In the formula, *accuracy* indicates oil price forecast accuracy under parameters of τ and m and γ and σ.

4.3. Unified solution Procedure of Oil Price Forecast Model Parameter

(1) The oil price time-series data are collected. Since LSSVM is most sensitive to the data between 0-1, normalized pretreatment of oil prices is performed, specifically:

$$\hat{x}_i = \frac{(x_i - x_{\min})}{(x_{\max} - x_{\min})} \tag{14}$$

In the formula, $\hat{x}_i$ is the normalized oil prices; xi is original oil prices; xmax and xmin are maximum and minimum values in original data.

Finally, anti-normalization processing of oil price predictions is conducted to restore the true predictive value, and the specific formula is as follows:

$$x_i = x_i \times (x_{\max} - x_{\min}) + x_{\min} \tag{15}$$

(2) Initial population is generated using random method, and each individual is composed of four parts of τ, m and γ, σ.

(3) The anti- coding of individual is performed, and reconstruction of oil price time-series is implemented according to τ and m. Learning of oil price time-series training set is conducted for LSSVM according to the parameters γ, σ to establish appropriate oil price forecasting model and to calculate fitness value for each individual.

(4) Whether the termination condition of algorithm is satisfied is determined, and if the termination condition is satisfied, then optimization is stopped, and the anti- coding of the best individual is performed to get the best τ, m and γ, σ. Then go to step (6).

(5) Genetic operations such as selection, crossover and mutation are performed for individual to generate the next generation of the population. Return to step (3) to continue the optimization of the parameters.

(6) Reconstruction of oil price time series is made in accordance with the optimal parameters τ and m, and learning of oil prices training set is performed for LSSVM using γ and σ to establish the optimal oil price forecasting model, and to predict test set, thus obtaining forecasted results of oil prices.

Establishing flow of oil price forecasting model is shown in Figure 1.

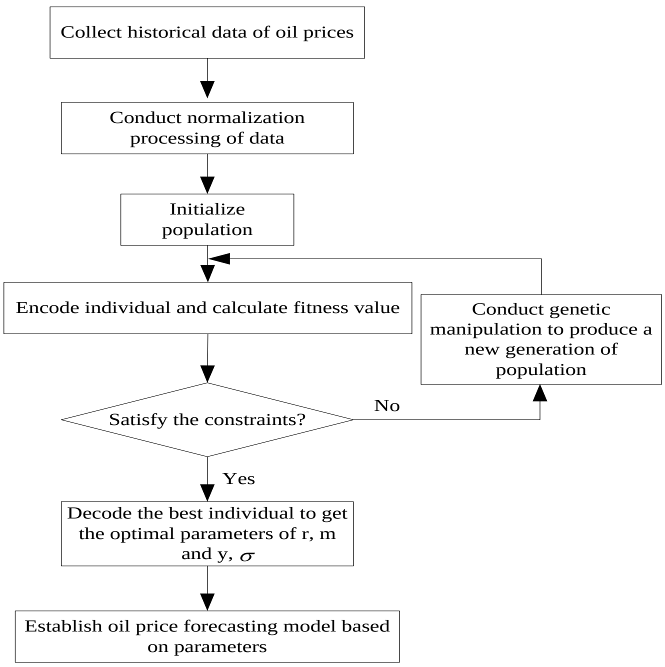

Figure 1. Establishing Flow of Oil Price Forecasting Model

5. Simulation Experiment

5.1. Data Sources

Daqing oil monthly prices on average from January 31, 2001 to April 30, 2013 are used as simulation objects, and a total of 160 data are collected, and the oil price time-series formed by them is shown in Figure 2. The former 100 data are taken as the training sets, while the later 60 data are regarded as test sets.

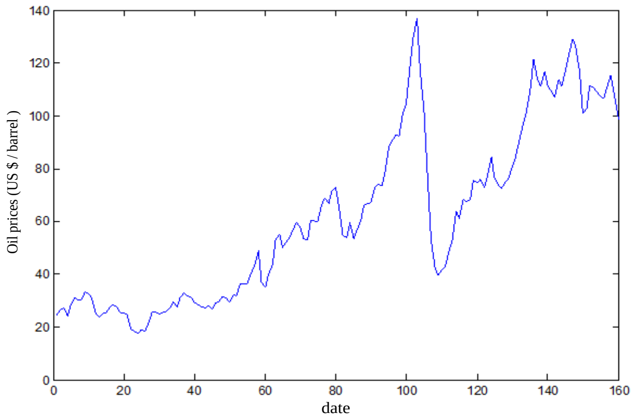

Figure 2. Oil Price Time-Series

5.2. Contrast Model and Evaluation Criteria

In order to make the forecasting results of PSR-LSSVM oil price model more comparable, two contrasting models are chosen as follows: ① oil price forecasting model (LSSVM) that mutual information method and false neighbor method are used to solve the point τ and m and GA is employed to optimize LSSVM parameters of γ and σ; ② oil price forecast model (CC-LSSVM) that CC is firstly used to solve τ and m, and then GA is employed to optimize LSSVM parameters of γ and σ. RMSE and MPAE are used as the criteria to evaluate pros and cons of forecasted results of the models.

5.3. Model Implementation

5.3.1. LSSVM Modeling Process: For the oil price time-series in Figure 2, the mutual information function is worked out using mutual information method, and the first minimum point is taken as the delay time. Figure 3 shows that when τ is 3, the first minimum value can be obtained, and τ is the time delay at this time.

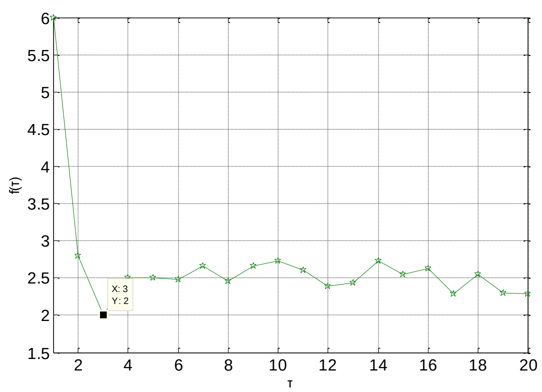

Figure 3. Time Delay Calculated by Mutual Information Method

Embedded dimension is calculated by false nearest neighbor method, and the results are shown in Figure 4, and the obtained embedding dimension is 4. The largest Lyapunov exponent of monthly average price time-series of Daqing oil calculated with a small amount of data is 1.163, which indicates that the monthly average price time-series of Daqing oil is chaotic.

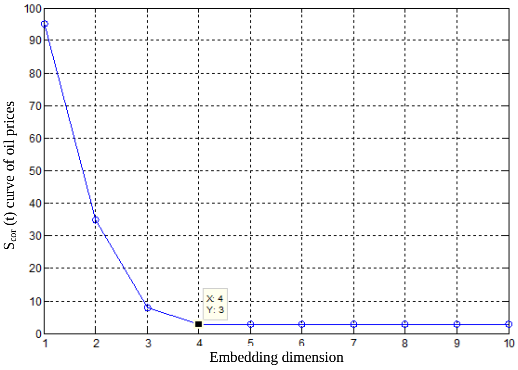

Figure 4. Embedding Dimension Calculated by FNN Method

Reconstruction of oil prices time-series is conducted on the basis of τ = 3 and m = 4, and then the reconstructed oil prices time series training set is input to LSSVM for learning, and γ and σ are optimized using GA to get the optimal γ = 20.12 and σ = 1.43, thereby establishing LSSVM-based oil price forecasting model.

5.3.2. CC-LSSVM Modeling Process: For the oil price time-series, CC method is used to conduct phase space reconstruction. Figure 5 is the calculation of the delay time τ using CC method, and it can be seen from the figure that the first approximate minimum value of $\Delta \bar{S}(t)$ is 4, and the first approximate zero of $\bar{S}(t)$ is also 4, thereby determining the delay time τ = 4 as the time delay of oil price time series.

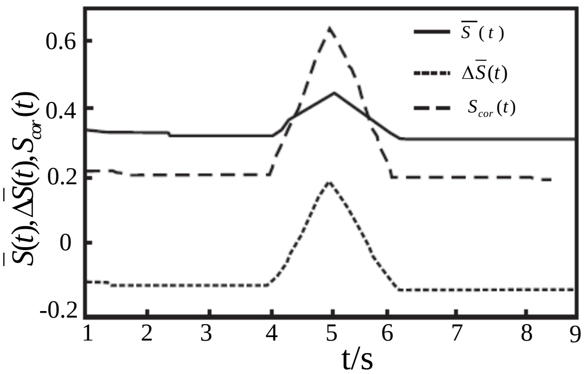

Figure 5. Time Delay of oil Prices Time-Series Calculated by CC Method

Figure 6 shows Scor (t) curve that the C-C method is used to calculate the oil prices time-series, and we can see that at around t = 18, the global minimum point can be achieved in Scor (t) on the basis of τw = (m-1) τ, thus taking m = 4.

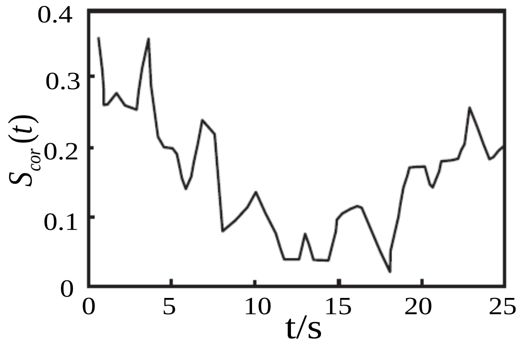

Figure 6. Scor (t) Curve of Oil Prices Time-Series Calculated by CC Method

Reconstruction of oil prices time-series is conducted on the basis of τ = 4 and m = 4, and then the reconstructed oil prices time series training set is input to LSSVM for learning, and γ and σ are optimized using GA to get the optimal γ = 147.55 and σ = 1.55, thereby establishing CC-LSSVM-based oil price forecasting model.

5.3.3. PSR-LSSVM Modeling Process: Based on the above-mentioned unified solving by parameter of oil price forecast model, the optimal parameters of PSR-LSSVM are obtained: τ = 1, m = 5, γ = 176.45 and σ = 6.22, and according to the optimal parameters, PSR-LSSVM-based oil prices predictive models are established, and the obtained three kinds of oil price forecasting model parameters are shown in Table 1.

Table 1. Oil Price Forecast Model Parameters

Prediction Model	*τ*	*m*	*γ*	*σ*
LSSVM	3	4	20.12	1.43
CC-LSSVM	4	4	147.55	1.55
PSR-LSSVM	1	5	176.45	6.22

5.4. Result and Analysis

5.4.1. Comparison in Generalization Ability of the Model: According to the established LSSVM, CC-LSSVM and PSR-LSSVM oil price forecasting models, fitting of oil prices training set is performed, and fitting results are shown in Figure 7.

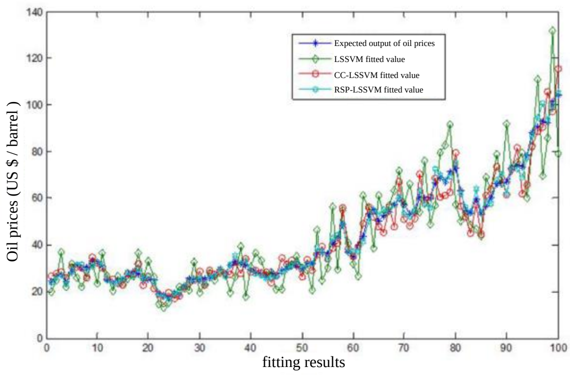

Figure 7. Oil Price Training Set Fitting Results for each Model

Figure 7 shows that among all forecasting models, fitting effect of PSR-LSSVM is at best, and its fitting results quite match the desired output of the oil price. This is mainly because the PSR-LSSVM has tapped the intrinsic link between phase space reconstruction and LSSVM through unified solution of τ and m parameters in phase space reconstruction and γ and σ parameters LSSVM, which can more accurately fit the complex trends of oil prices to further improve the fitting precision of oil prices. Comparison results show that the idea of PSR-LSSVM used for modeling of the oil price is feasible and effective, and its performance is superior to that of oil price forecasting models of τ and m and γ and σ through traditional separation and independent optimization.

5.4.2. Comparison in Generalization Ability of the Model: Evaluation of the pros and cons of predictive model performance mainly focuses on its ability to predictive capacity rather than fitting ability. Therefore, LSSVM, CC-LSSVM and PSR-LSSVM oil price forecasting models make prediction for the test set respectively, forecasted results shown in Figure 8. From Figure 8, the prediction performance of PSR-LSSVM is better than the comparison model, and the predicted results are more consistent with the actual value of oil price, thus it is an oil price forecasting model of high prediction accuracy and strong generalization ability.

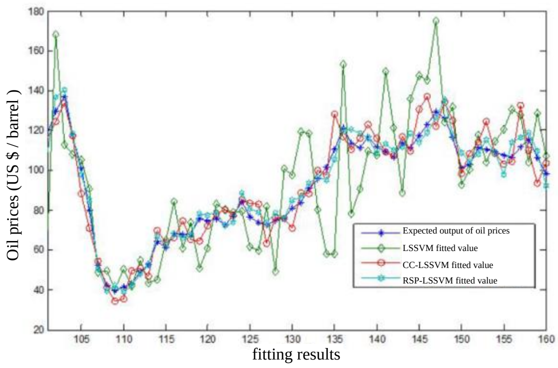

Figure 8. Forecasted Results of Oil Prices Test Set for Each Model

Prediction fitting and prediction error of each model are shown in Table 2. From various evaluation criteria in Table 2, PSR-LSSVM prediction error is much smaller than that of LSSVM and CC-LSSVM, and the prediction accuracy can be improved. A comprehensive comparison of the results shows that PSR-LSSVM performing unified solution of phase space reconstruction and prediction algorithm parameters can improve the prediction accuracy of oil prices to overcome the defects difficult to find the global optimum parameters through traditional separation and independent optimization, with more reliable predictions.

Table 2. Performance Comparison of Different Oil Prices Prediction Models

Prediction model	Training set		Test set	
	RMSE	*MAPE*	*RMSE*	*MAPE*
LSSVM	9.691	16.27%	21.204	17.61%
CC-LSSVM	4.534	6.30%	7.115	7.21%
PSR-LSSVM	2.043	3.39%	4.279	3.81%

6. Conclusion

Oil prices are time-varying and chaotic, and it needs to reconstruct phase space and optimize parameters of prediction algorithm in the modeling and forecasting process of oil prices. Therefore, the links between them are fully used to dig out the complex trends of oil prices, and a PSR-LSSVM-based oil price forecast model is proposed. Simulation results show that relative to the comparison model, PSR-LSSVM has raised the oil price forecast accuracy, and research results have important theoretical and practical significance to modeling and prediction for chaotic oil prices.

Acknowledgments

This work was supported by the Hunan province soft science research plan Key projects (2014ZK2041).

References

[1] J. He, Y. Geng, Y.Wan, S. Li and K. Pahlavan, “A cyber physical test-bed for virtualization of RF access environment for body sensor network”, IEEE Sensor Journal, vol. 13, no. 10, Oct. **(2013)**, pp. 3826-3836,
[2] W. Huang, Y.Geng, Identification Method of Attack Path Based on Immune Intrusion Detection, Journal of Networks, vol. 9, no. 4, Jan. **(2014)**, pp. 964-971.
[3] Y. Su, “In-situ bitmaps generation and efficient data analysis based on bitmaps”, In Proceedings of the 24th International Symposium on High-Performance Parallel and Distributed Computing. ACM, **(2015)**., pp. 61-72.
[4] G. Yan, Y. Lv, Q. Wang and Y. Geng, “Routing algorithm based on delay rate in wireless cognitive radio network”, Journal of Networks, vol. 9, no. 4, Jan. **(2014)**, pp. 948-955.
[5] D. Jiang, X. Ying, Y.Han and Z. Lv. “Collaborative Multi-hop Routing in Cognitive Wireless Networks”, Wireless Personal Communications, **(2015)**.
[6] X. Zhou, "Exploring Distributed Resource Allocation Techniques in the SLURM Job Management System", Illinois Institute of Technology, Department of Computer Science, Technical Report, **(2013)**.
[7] G. Bao, L. Mi, Y. Geng, M. Zhou and K. Pahlavan, “A video-based speed estimation technique for localizing the wireless capsule endoscope inside gastrointestinal tract”, 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Aug. **(2014).**
[8] D. Zeng and Y. Geng, “Content distribution mechanism in mobile P2P network”, Journal of Networks, vol. 9, no. 5, Jan. **(2014)**, pp. 1229-1236,
[9] T. Li, "ZHT: A light-weight reliable persistent dynamic scalable zero-hop distributed hash table", Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on. IEEE, **(2013)**.
[10] K. Wang, "Optimizing load balancing and data-locality with data-aware scheduling." Big Data (Big Data), 2014 IEEE International Conference on. IEEE, **(2014)**.
[11] M. Zhou, G. Bao, Y. Geng, B.Alkandari and X. Li, “Polyp detection and radius measurement in small intestine using video capsule endoscopy”, 2014 7th International Conference on Biomedical Engineering and Informatics (BMEI), Oct. **(2014)**
[12] X. Song and Y. Geng, “Distributed community detection optimization algorithm for complex networks”, Journal of Networks, vol. 9, no. 10, Jan. **(2014)**, pp. 2758-2765.
[13] J. Yang, S. He, Y. Lin and Z. Lv, “Multimedia cloud transmission and storage system based on internet of things. Multimedia Tools and Applications”, **(2016)**.
[14] J. Yang, B. Chen, J. Zhou and Z. Lv, “A portable biomedical device for respiratory monitoring with a stable power source”, Sensors, **(2015)**.
[15] D. Zhao, "FusionFS: Toward supporting data-intensive scientific applications on extreme-scale high-performance computing systems." Big Data (Big Data), 2014 IEEE International Conference on. IEEE, **(2014)**.
[16] S. Dang, J. Ju, D Matthews, X. Feng and C. Zuo, “Efficient solar power heating system based on lenticular condensation”, Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference, April **(2014)**, pp. 26-28

[17] X. Zhang, Y. Han, D. Hao and Z. Lv, "ARPPS：Augmented Reality Pipeline Prospect System", 22th International Conference on Neural Information Processing Istanbul, Turkey, In press, ICONIP (**2015**).
[18] J. Hu and Z. Gao, "Distinction immune genes of hepatitis-induced heptatocellular carcinoma", [J] Bioinformatics, vol. 28, no.24, **(2012)**, pp. 3191-3194

Author

Hu Shuyu received her M.S. degree in software engineering from Hunan university in Changsha, China. She is currently a lecturer in Hunan radio and tv university. Her research interest is mainly in the area of Computer Software, Mechanical and Electrical Integration, Management of the Economy. She has published several research papers in scholarly journals in the above research areas and has participated in several books.