Reference

1. Hua Li, Hyun S. Yang. Fast and reliable image enhancement using fuzzy relaxation technique. IEEE Transactions on Systems, Man and Cybernetics, 2011, 19 (5):1276~1281.
2. Dongling Peng, Tie-un Wu. A generalized image enhancement algorithm using fuzzy sets and its application. Proceedings of the 1st International Conference on Machine Learning and Cybemetics, 2012, 820~823.
3. H.R. Tizhoosh, G. Krell, B. Michaelis. On fuzzy enhancement of megavoltage images in radiation therapy. Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, 2007, 3: 1398~1404.
4. S. K. Pal, R. A. King. Image enhancement using smoothing with fuzzy sets. IEEE Transactions on Systems, Man and Cybernetics, 2011, 11(7):494~501.
5. S. K. Pal, R. A. King. On edge detection of X-ray images using fuzzy sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010,5(1):69~77
6. Zheng Chunhong, Jiao Licheng, Chen Xiping, Yuan Zhanting. A new fast fuzzy processing method for B-scan image. Proceedings of the 4th World Congress on Intelligent Control and Automation, 2002: 6~9.
7. Pan Rongjian, Meng Xiangxu. A method of local enhancement based on fuzzy set theory. Proceedings of 3D World Congress on Intelligent Control and Automation, 2000, 1751~1753.
8. Wen Bingtao, Jin Wentian, Jian Liu. Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm. Pattern Recognition Letters, 2003, 24: 3069~3078
9. Li Xiangji, Ding Runtao. Fuzzy morphological operators to edge enhancing of images. Proceedings of International Conference on Signal Processing, 2008, 2: 1017~1020
10. Fa-Shen Leou, Kuei-Ann Wen. Image enhancement based on the visual model using the concept of fuzzy set. Proceedings of IEEE International Symposium on Circuits and Systems, 2011, 5: 2581~2584.
11. Liu Jinhua, Tang Jingxin, Long Tujing. An improved fast algorithm for fuzzy edge detection. System simulation, 2003, 15 (2): 273~274
12. Sun Wei, Xia Liangzheng, Pan Han. An edge detection algorithm based on fuzzy division. China Image and Graphics, 2004, 9 (1): 18~22

FCM-LSSVM Based on Training Sample Selection

Niu Xuetong

International School of Beijing University of Posts and Telecommunications, Beijing 102209, China

Corresponding author is Niu Xuetong

Abstract

To improve the predictive accuracy of network traffic, as for the training sample selection, this paper proposes a FCM-LSSVM for training sample selection. Firstly, fuzzy-means clustering algorithm is used to make cluster analysis of network filling data to eliminate the isolated sample points in it and to build training set of LSSVM, and

then the training set is input to the LSSVM for learning. Furthermore, Artificial Bee Colony is used to optimize model parameters, and finally, network traffic forecasting model is established, and the simulation experiment is employed to test the performance of the model. Simulation results show that compared to other network traffic forecasting models, FCM-LSSVM not only improves the prediction accuracy of network traffic, and increases modeling speed, getting more desirable forecasted results of network traffic.
Keywords: NETWORK TRAFFIC, LSSVM, FUZZY MEANS CLUSTERING, TRAINING SET

1. Introduction

With the development of network technology, there is a wide range of network services and more frequent network congestion, and network traffic prediction is one of the key technologies of network management, but the network traffic is affected by multiple factors, such as holidays, price, online behavior, and so on, and its change is periodic and similar, as well as mutant and non-stationary, thus the establishment of network traffic prediction model with high accuracy has been a hot issue concerned by domestic and foreign experts and scholars[1].

For complicated and variable network traffic, numerous effective network traffic models arise. Traditional network traffic prediction models: Markov process, multiple linear regression, Poisson process, time sequences, etc. [2,3], because network traffic data is collected in chronological order, a typical time-series data, time series prediction method is widely used in main fields such as: ARMA, FARIMA, and exponential smoothing, etc. Time series is a kind of linear modeling method, only able to conduct accurate prediction+n of short-range dependence network traffic, but it is difficult to establish an accurate model for the long--range dependence network traffic with low prediction accuracy, and the results are unreliable [4]. As the nonlinear theory continues to mature, there has been network traffic prediction model based on neural networks, and compared to the traditional model, it can make accurate prediction for the non-stationary change characteristics of network traffic, and the prediction accuracy of network traffic can be improved [5,6]. However, due to its defects of easily occurred close fitting, and poor ability of generalization, they lead to be easy to get local optimal network traffic prediction. Based on modern statistical theory, support vector machine (SVM) fusing structural risk minimization principle and nuclear technology better overcomes the deficiencies of neural networks with excellent abilities of generalization, and obtains more ideal forecasted results of network traffic [7]. But for large-scale network traffic, there exists slow speed of training and longer time in SVM difficult to meet the online prediction requirements of network traffic. LSSVM has improved SVM and simplified operation, with a small amount of computing, which accelerates the speed of modeling network traffic, able to obtain online prediction model requirements for network traffic [8]. In practice, when training samples contain isolated sample, these samples increase training time of LSSVM, but also have an adverse effect on the predicted results, and therefore the accurate selection of isolated samples in network traffic samples and the elimination of the adverse effects on samples are of great significance to improve the prediction accuracy of network traffic [9].

In order to improve the prediction accuracy of network traffic, the FCM-LSSVM for training sample selection is proposed. Firstly, FCM is used to eliminate outliers, to reduce training sample size, and to build training samples of network traffic, which has improved the generalization performance of LSSVM, and then LSSVM is adopted to establish networking traffic prediction model based on the training sets, and finally the performance of the model is tested through simulation experiment. The simulation results show that, FCM-LSSVM only improves the prediction accuracy of network traffic, while accelerating the speed of modeling network traffic.

2. LSSVM and FCM

2.1. LSSVM

LSSVM is an improved support vector machine (SVM), and it transforms inequality constraints in SVM into equality constraints, using the least-squares linear system as a loss function. Network traffic forecasting is a complex and high dimensional nonlinear problem with limited samples, and therefore, LSSVM is used to describe this non-linear relationship. The basic principle of this method is described as follows: the sample of training data can be expressed as $\{x_i, y_i\}_{i=1}^{l}$: $x_i \in R^n$ is the input vector of the i-th samples; $y_i \in R^n$ is the target value of the i-th sample, and l *is* the number of training samples. The following form of LSSVM model is taken in the feature space

$$f(x) = \omega^T \cdot \varphi(x) + b \tag{1}$$

In the formula, $\varphi(x)$ means mapping the input sample data into high dimensional feature space; ω indicates weight vector.

According to the structural risk minimization principle, through comprehensive consideration of the complexity of function and fitting error, objective function of LSSVM can be described as

$$\min_{\omega,b,\xi} J(\omega,\xi)=\frac{1}{2}\omega^T\omega+\frac{\gamma}{2}\sum_{i=1}^{n}\xi_i^2$$

$$s.t. \qquad (2)$$

$$\begin{cases} y_i\left[\omega^T\varphi(x)+b\right]=1-\xi_i \\ i=1,2,n,,l \end{cases}$$

In the formula, ξ_i is the error variance; γ is adjustable parameter, used for controlling the punishment level of sample exceeding the error.

Obviously, LSSVM only has equality constraints, and the loss function in the target is the two-norm of error, which greatly simplifies the problem solving and Lagrange function is defined as

$$L(\omega,b,\xi,\alpha)=J(\omega,\xi)-\sum_{i=1}^{l}\left(\alpha_i\{y_i[\omega^T\varphi(x_i)-b]\}-1+\xi_i\right) \quad (3)$$

In the formula, α_i is the Lagrange multiplier.

According optimal conditions of Karush-Kuhn-Tucker (KKT), $\frac{\partial L}{\partial\omega}=0$, $\frac{\partial L}{\partial b}=0$, $\frac{\partial L}{\partial\xi_i}=0$, $\frac{\partial L}{\partial\alpha_i}=0$ i = 1,2, ..., l are calculated successively, and after elimination of ξ_i and ω, the following linear equations are obtained

$$\begin{bmatrix} 0 & Q^T \\ Q & PP^T+\gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (4)$$

Among them, $P=[\varphi(x_1)^T y_1 n\varphi(x_l)^T y_l$; $I=[1n1]^T$; $Q=[y_1 ny_l]^T$; $a=[a_1 na_l]^T$.

According to Mercer condition, core function can be written as

$$k(x_k,x_h)=\varphi(x_k)^T\varphi(x_h) \quad (5)$$

The formula (4) can be modified to

$$\begin{bmatrix} 0 & y_1 & n & y_l \\ y_1 & y_1y_1k(x_1,x_1)+\gamma^{-1} & n & y_1y_1k(x_1,x_l) \\ n & n & n & n \\ y_1 & y_1y_1k(x_l,x_l)l & n & y_ly_lk(x_l,x_l)+\gamma^{-1} \end{bmatrix}\begin{bmatrix} b \\ a_1 \\ n \\ a \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \\ n \\ 1 \end{bmatrix} \quad (6)$$

Obviously, optimization problems of LSSVM are converted into the solving of above linear equations. LSSVM usually has faster training speed than the standard support vector machine. After b and α, in the formula (6) are figured out, regression function of LSSSVM is

$$y(x)=\sum_{i=1}^{l}\alpha_i k(x,x_i)+b \quad (7)$$

For different kernel functions, different FCM-LSSVMs can be created, because the radial basis kernel function has advantages of fewer parameters and good universality, etc., paper chooses it as LSSVM kernel function, defined as follows:

$$k(x,x_i)=\exp\left(-\frac{|x-x_i|^2}{\sigma^2}\right) \quad (8)$$

2.2. FCM

Provided that data sample X = {x1, x2, ..., xn}, n represents the number of samples, and cost function of FCM algorithm is:

$$\min J_m(\boldsymbol{U},\boldsymbol{Z})=\sum_{j=1}^{N}\sum_{i=1}^{C}(u_{ij})^m d_{ij}^2 \quad (9)$$

Corresponding constraints are

$$\begin{cases} \sum_{k=1}^{c}u_{ik}=1 \\ u_{ik}\in[0,1] \\ \sum_{i=1}^{n}u_{ik}>0 \end{cases} \quad (10)$$

In the formula, C represents the number of clusters; uij is the membership degree of sample xj to the ith class; zi is the cluster center of the ith class; dij = || xj-zi || is measure of similarity.

3. Network Traffic Predicting Model Based on FCM-LSSVM

In the process of collecting network traffic, due to various factors, network traffic data inevitably contains noise data, that is, isolated samples, and this paper using the FCM conducts clustering analysis of network traffic to select important network traffic, and then LSSVM is used to establish networking traffic forecast model. Specific steps are as follows:

(1) History data of network traffic are collected, and they are divided into two types of network traffic data, and the number of cluster centers m of traffic data is given, and processing is conducted for network traffic according to formula (11) to eliminate

too large or too little data range so as to avoid adverse impact on the training process.

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (11)$$

In the formula, x is the raw data of network traffic, and xmax and xmin are upper and lower bounds respectively.

(2) The membership matrix is computed according to formula (12)

$$u_{ij} = \begin{cases} 1 / \sum_{k=1}^{C} (\frac{d_{ij}}{d_{ik}})^{\frac{2}{m-1}} & , d_{ij} \neq 0 \\ 1 & , d_{ij} = 0, j = k \\ 0 & , d_{ij} = 0, j \neq k \end{cases} \quad (12)$$

Among it, n is the total sample set.

(3) The cost function J is calculated, if J <R, in which R represents the threshold, then it means stop to get the final cluster center C, fuzzy membership matrix U and distance matrix D, go to step (5), or go to step (4).

(4) The cluster centers Ci is recalculated according to equation (13).

$$c_i = \sum_{j=1}^{n} u_{ij}^m x_j / \sum_{j=1}^{n} u_{ij}^m \quad (13)$$

(5) The category of the sample is determined on the basis of KNearest Neighbor to select training set of LSSVM and to eliminate network traffic sample of outliers.

(6) The network traffic training set is input into LSSVM for training, and Artificial Bee Colony is used to find the optimal parameters of LSSVM and to establish the optimal network traffic prediction model.

(7) The forecasting model is employed to predict the test of network traffic and to analyze the forecasted results.

Based on the above mentioned, workflow of network traffic forecasting model based on FCM-LSSVM is shown in Figure 1.

4. Simulation Experiment

4.1 Network Traffic Data

To test the performance of network traffic model proposed, this paper selects network traffic library: http: //newsfeed.ntcu.net/~news/2013/, collecting the visiting traffic per hour in everyday network from October 1, 2013 to October 30 of incoming articles on host node router to obtain 720 data, specifically shown in Figure 2. The former 620 samples are used to train network traffic prediction model, and to establish a network traffic prediction model, while the last 100 samples are tested as the testing of generalization ability of model on test set.

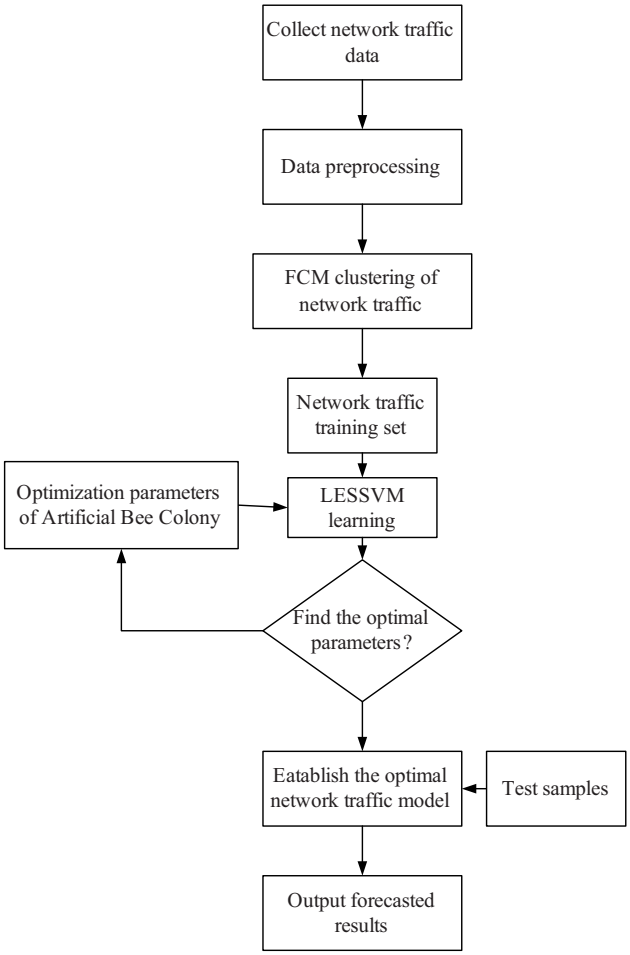

Figure 1. Workflow of Network Traffic Flow Forecasting Model

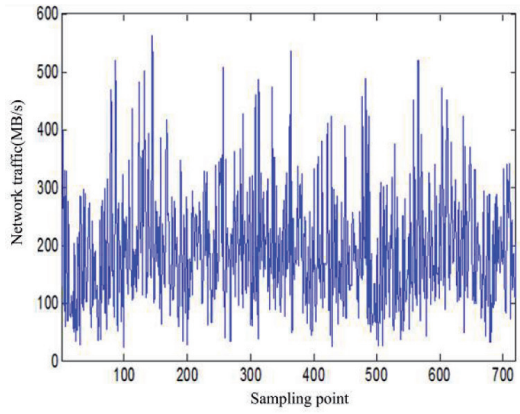

Figure 2. Network Traffic Data

4.2 Contrast Model and Performance Evaluation Criteria

Comparative experiments are performed in LSSVM without cluster analysis, FCM-RBFNN, and FCM-BPNN and FCM-SVM, and the performance evaluation index of model is: root mean square error (RMSE) and mean relative error (MAPE), which are defined as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|^2} \quad (15)$$

In the formula, n represents the number of test set samples; y_i and $\hat{y}_i$ are the actual value and the predictive value of the network traffic respectively.

4.3 Result and Analysis

Firstly, clustering analysis of network traffic in FIG. 1 is made using FCM to choose the most effective training samples and to delete the isolated sample points, and then LSSVM is used to establish a network traffic prediction model, and forecasted results and forecast errors of test set for FCM-LSSVM and contrast model are shown in Figures 3 and 4. From Figure 3, the forecasted results of FCM-LSSVM are fairly consistent with the value of actual network traffic and are very close to one another, suggesting that FCM-LSSVM can be used for modern complex and random network traffic prediction, but as can be seen from Figure 4, the prediction error of FCM-LSSVM is quite small, which indicates high prediction accuracy of FCM-LSSVM, able to meet network traffic prediction accuracy.

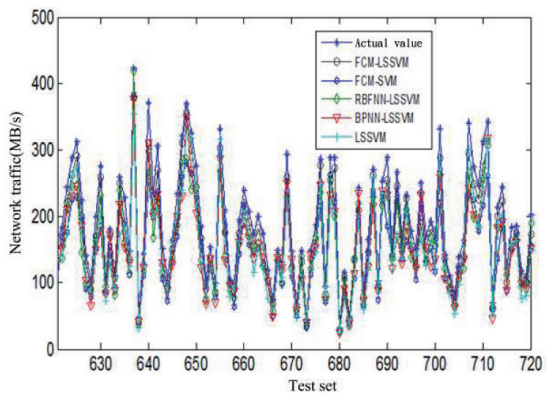

Figure 3. Predictions for the Training Set in Each Model

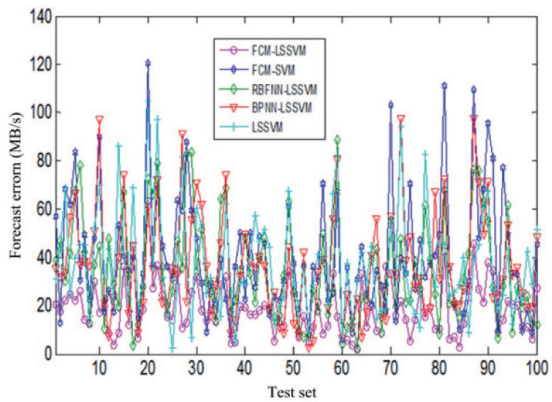

Figure 4. Error Comparison of the Prediction for the Training Set in Each Model

Through further detailed analysis of Figure 3 and Figure 4, the following conclusions can be drawn:

(1) Compared to LSSVM without cluster analysis, prediction accuracy of FCM-LSSVM can be improved, reducing the average error of prediction, suggesting that using FCM algorithm to make clustering analysis of the network traffic samples can eliminate the adverse effects of isolated samples on the modeling process, getting better optimal training sample set, enhancing sample data regularity, and establishing a network traffic forecasting model with higher prediction accuracy, which indicates that the proposed network traffic modeling thought is feasible.

(2) Compared to FCM-SVM, FCM-LSSVM can get even better predictions, suggesting that LSSVM has improved SVM to speed up the learning speed of network traffic and to establish a network traffic prediction model with higher overall forecast performance, which has better fitted network traffic trends, and has increased network traffic prediction accuracy.

(3) The prediction accuracy of FCM-LSSVM has been greatly improved relative to FCM-BPNN, FCM-RBFNN, which is mainly because LSSVM has better overcome BPNN and RBFNN, and based on modeling principle of empirical risk minimization, it has avoided deficiencies of excessive fitting, local optimum, conducting the most accurate portrayal of network traffic change regularity, thereby achieving network traffic forecast results with higher accuracy.

Prediction errors of FCM-LSSVM and comparison model are shown in Table 1. According to Table 1, compared to the comparison model, FCM-LSSVM has greater prediction error, and the corresponding network traffic forecasting accuracy is also higher, which indicates that FCM-LSSVM selects sample of network traffic by FCM, eliminating the defect that LSSVM is sensitive to training samples, while modeling is conducted using LSSVM to overcome defects of low prediction accuracy for SVM, neural networks, etc. FCM-LSSVM is a network traffic flow forecasting model with high precision and reliable forecasted results.

Table 1. Performance Comparison of Different Network Traffic Prediction Models

Prediction Model	*RMSE*	*MAPE*(%)
LSSVM	12.33	10.12
FCM-SVM	9.90	6.23
FCM-RBFNN	11.57	7.90
FCM-BFNN	12.17	8.12
FCM-LSSVM	4.01	3.05

4.4. Analysis of Noisy Network Traffic Forecasted Results Based on FCM-LSSVM

To test the versatility of FCM-LSSVM, white noise is added to network traffic in Figure 1 to obtain noisy network traffic, and the specific is shown in Figure 5. Forecasted results and prediction errors of noisy network traffic data based on FCM-LSSVM are shown in Fig. 6 and Fig.7. From Fig. 6 and Fig.7, as for the noisy network traffic, FCM-LSSVM has also received ideal forecast results, which shows that FCM-LSSVM employs FCM to pretreat network traffic, and network traffic with high noise removed as isolated sample, obtaining a better network traffic training set, increasing the robustness and versatility of the model, thereby broadening the applications of FCM-LSSVM in network traffic.

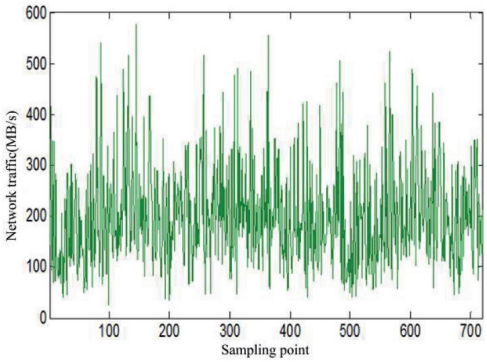

Figure 5. Noisy Network Traffic Data

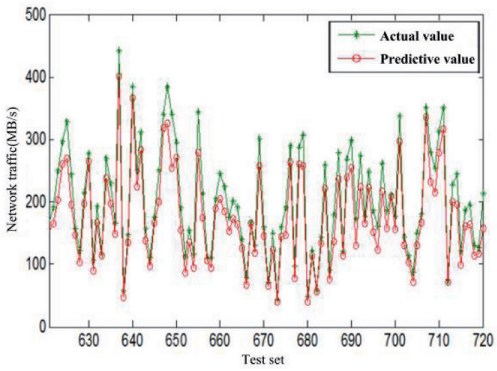

Figure 6. Analysis of Network Traffic Forecasted Results Based on FCM-LSSVM

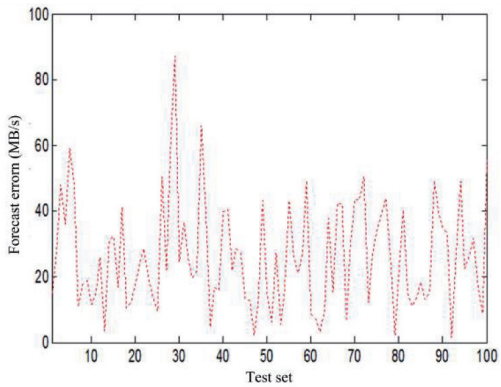

Figure 7. Prediction Error of Noisy Network Traffic Based on FCM-LSSVM

Noisy network traffic prediction errors of FCM-LSSVM and contrast model are shown in Table 2. From Table 2, compared to the comparison model, the overall forecast performance of FCM-LSSVM is better, and network traffic forecasting model of smaller error is established, which has proved the superiority of FCM-LSSVM once again.

Table 2. Comparison of Noisy Network Traffic Prediction Errors in Different Models

Prediction Model	*RMSE*	*MAPE*
LSSVM	20.14	15.10
FCM-SVM	16.18	13.68
FCM-RBFNN	18.00	15.19
FCM-BFNN	17.76	14.72
FCM-LSSVM	9.24	7.23%

Conclusions

In order to improve the prediction accuracy of network traffic, because isolated samples in network traffic data have adverse effects on predictive performance of LSSVM, this paper proposes FCM-LSSVM based on training sample selection, implementing clustering of network traffic data through FCM to choose the most effective training data and to delete outliers. Simulation results show that, FCM-LSSVM not only improves the prediction accuracy of network traffic, while accelerating the speed of network traffic forecast, thus providing a new research idea for the complex network traffic.

References

1. Yu Song, et al.. A Rapid and High Reliable Identify Program for Nighttime Pedestrians. Infrared Physics & Technology, 2015,27,pp.58-60.
2. Dingde Jiang, Zhengzheng Xu, Wenpan Li, Zhenhua Chen. Network Coding-Based Energy-Efficient Multicast Routing Algorithm for Multi-hop Wireless Networks. Journal of Systems and Software, 2015,12,pp.125-128.
3. Yishuang Geng, Jin Chen, Ruijun Fu, Guanqun Bao, Kaveh Pahlavan. Enlighten wearable physiological monitoring systems: On-body rf characteristics based human motion classification using a support vector machine. IEEE transactions on mobile computing, 2015,4,pp.12-15.
4. Jie He, Yishuang Geng, Fei Liu, Cheng Xu. CC-KF: Enhanced TOA Performance in Multipath and NLOS Indoor Extreme Environment. IEEE Sensor Journal, 2014,14,pp. 3766-3774.

5. Shuang Zhou, Liang Mi, Hao Chen, Yishuang Geng. Building detection in Digital surface model, 2013 IEEE International Conference on Imaging Systems and Techniques (IST), NewYork,2013,pp.88-91
6. Jie He, Yishuang Geng, Kaveh Pahlavan. Toward Accurate Human Tracking: Modeling Time-of-Arrival for Wireless Wearable Sensors in Multipath Environment. IEEE Sensor Journal, 2014,14, pp.3996-4006.
7. Na Lu, Caiwu Lu, Zhen Yang, Yishuang Geng. Modeling Framework for Mining Lifecycle Management. Journal of Networks, 2014,9, pp.719-725.
8. Guanxiong Liu, Yishuang Geng, Kaveh Pahlavan. Effects of calibration RFID tags on performance of inertial navigation in indoor environment, 2015 International Conference on Computing, Networking and Communications (ICNC),London, 2015,pp.22-25.
9. Jie He, Yishuang Geng, Yadong Wan, Shen Li, Kaveh Pahlavan. A cyber physical test-bed for virtualization of RF access environment for body sensor network. IEEE Sensor Journal, 2013,10, pp.3826-3836.
10. Jiachen, Yang, et al.. Objective Evaluation Criteria for Stereo Camera Shooting Quality under Different Shooting Parameters and Shooting Distances. Sensors Journal,2012, 11,PP.147-149.
11. Zhihan Lv. Wearable Smartphone: Wearable Hybrid Framework for Hand and Foot Gesture Interaction on Smartphone, IEEE International Workshop on Wearable Computer Vision Systems at International Conference on Computer Vision 2013 (ICCV 2013), Sydney, 2013, 12,pp.1258-1261.
12. Dingde Jiang, Zhengzheng Xu, Peng Zhang, and Ting Zhu. A transform domain-based anomaly detection approach to network-wide traffic. Journal of Network and Computer Applications, 2014,40,pp. 292-306.
13. Xiaoming Li, Zhihan Lv, Baoyun Zhang, Weixi Wang, Shengzhong Feng, Jinxing Hu. XEarth: A 3D GIS Platform for managing massive city information. 2015 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications(CIVEMSA 2015),NewYork,2015,7,pp.98-102.
14. Liguo Zhang, Binghang He, Jianguo Sun, Mingzhu Lai, Zhihan Lv. Double Image Multi-Encryption Algorithm based on Fractional Chaotic Time Series. Journal of Computational and Theoretical Nanoscience, 2014,42,pp.63-66.
15. Wei Luo, Zhiyong Wang, Zhihan LV. Method to Acquire a Complete Road Network in High-resolution Remote Sensing Image Based on Tensor Voting Algorithm. EXCLI JOURNAL, 2015, 14,pp.123-125.
16. Jiachen, Yang, et al.. Multiview image rectification algorithm for parallel camera arrays. Electron. Imaging, 2013,23,pp.303-305.
17. Alex Tek, et al. Advances in Human-Protein Interaction-Interactive and Immersive Molecular Simulations. Biochemistry, Genetics and Molecular Biology.Protein-Protein Interactions-Computational and Experimental Tools, 2012,32,pp.27-65.
18. Tianyun Su, et al.. Rapid Delaunay Triangulation for Random Distributed Point Cloud Data Using Adaptive Hilbert Curve. Computers & Graphics , 2015,45,pp.542-545.
19. Xiaoming Li, et al.. WebVRGIS Based Traffic Analysis and Visualization System. Advances in Engineering Software ,2015,67,pp.147-151.