Acta Technica *62* (2017), No. 7A, 137–148 © 2017 Institute of Thermomechanics CAS, v.v.i.

Intrusion detection system based on hypergraph feature reduction and convolutional neural network

Tiancheng Wang[1]

Abstract. Since computer network detection is concealed and diverse, it is difficult to carry out network intrusion detection. In order to improve the detection effect of LSSVM in the abnormal network intrusion detection, an improved particle swarm optimization is introduced to optimize the network intrusion detection, and a network intrusion detection model integrating improved particle swarm optimization algorithm (IPSO) with least squares support vector machine (LSSVM) is proposed. Of which, the least squares support vector machine is used to replace support vector machine to reduce the complexity of algorithm to improve computing speed, where radial basis function is selected as the kernel function, and the standard parameters of normalization parameter and kernel function are regarded as two particles, and particle swarm optimization is used to optimize and seek the global optimum value. Aimed at the premature convergence defect of particle swarm optimization, the method of controlling population diversity is used to improve algorithm performance through initial population selection and premature convergence judgment. The test result of using KDDCUP99 DataSet shows the application effect of the improved particle swarm optimization to least squares support vector machine model (IPSO-LSSVM) in network intrusion detection.

Key words. Network intrusion detection, Neural network, Least squares support vector ma-chine, Population diversity, Improved particle swarm optimization.

1. Introduction

Intrusion detection refers to the analysis of network system logs and traffic data to determine whether there is the behavior harmful to the safety of computer systems [1]. With computer network is increasingly applied to in the work and life, more and more people become focused on the network security issue. And as a new generation of networks security technology, intrusion detection technology shows an increasingly strong effect in protecting network information security [2]. At present,

[1] School of Information Science and Engineering, Changzhou University, Changzhou, Jiangsu, 213164, China

the basic principles of most intrusion detection technologies is pattern matching, that's, matching the detected data information to the data information of rule base, and such method shows relatively high accuracy and low false alarm rate on the known network intrusion detection, but it cannot carry out effective recognition on the unknown intrusion or the variety of known intrusion [3], and the other shortcoming of pattern-matching detection is that a relatively complete rule base and knowledge base shall be established before the detection, followed by constant upgrade and maintenance of the base according to the emerging network intrusion mode; otherwise, the false alarm rate of detection will be increased continuously. Aiming at the common four kinds of attach types at present, the popular artificial neural network detection method shows a relatively detection rate of only about 70%, far to meet the security protection requirements of network [4]. However, the network intrusion detection technology based on abnormality can detect unknown network attack, and carry out pre-alarm to some extent on the security event, and now it has hence become the focus of current research [5].

2. IPSO-LSSVM model

2.1. LSSVM model

SVM model [10] performs data separation through a hyperplane determined by a certain number of support vectors, maps the data to be separated to the high-dimensional feature space through kernel function and in which linear hyperplane is used for data separation. Suppose sample data as $\{(x_1, y_1), \cdots, (x_i, y_i), \cdots (x_n, y_n)\}$, and its sample capacity is n, sample input value is $x_i \in R^d$ and $y_i \in$ R is the corresponding sample output, the optimal separating hyperplan is obtained as the equation (1).

$$f(x) = \omega^T \bullet \phi(x) + b = 0\,. \tag{1}$$

Where, ω is weight vector, and b is threshold value. Suppose the structural risk minimization condition as:

$$R = \|\omega\|^2 \Big/ 2 + c\sum_{i=1}^{n} \xi_i^2\,. \tag{2}$$

And the equation (3) shall be met:

$$y_i = \omega^T \phi(x_i) + b + \xi_i,\ i = 1, 2, \cdots, n\,. \tag{3}$$

Where, $\xi_i \in \xi$ represents the forecast error vector of training set, and its corresponding lagrangian function is:

$$L(\omega, b, \xi, \alpha) = R - \sum_{i=1}^{n} \alpha_i \left(\omega^T \bullet \phi(x_i) + b + \xi_i - y_i\right). \tag{4}$$

Where $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_n]$ is the lagrangian multiplier, and according to the optimization condition, there should be a partial derivative of 0, namely

$$\frac{\partial L}{\partial \omega} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial \xi_i} = 0, \frac{\partial L}{\partial \alpha_i} = 0\,. \tag{5}$$

Its solution can be got as:

$$.\begin{cases} \omega = \sum_{i=1}^{n} \alpha_i \phi(x_i) \\ \sum_{i=1}^{n} \alpha_i = 0 \\ 2c\xi_i = \alpha_i \end{cases} \tag{6}$$

Where c is a normalization parameter, and after combining equation (3) and equation (6), it can be got:

$$\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & K(x_1,x_2) + \frac{1}{2c} & K(x_1,x_2) & \cdots & K(x_1,x_2) \\ 1 & K(x_2,x_2) & K(x_2,x_2) + \frac{1}{2c} & \cdots & K(x_2,x_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_n,x_1) & K(x_n,x_2) & \cdots & K(x_n,x_n) + \frac{1}{2c} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} b \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \tag{7}$$

Where $K(x_i, x_j) =< \phi(x_i), \phi(x_j) >$ is a kernel function that meets Mercer symmetric function condition maps input space to high-dimensional feature space in a nonlinear way. At present, the common kernel functions include polynomial kernel function, Sigmoid kernel function, radial basis function and linear kernel function, etc.; based on the good features of radial basis function, radial basis function is selected in the Paper to use as the classification model of algorithm, and the corresponding expression is below:

$$K(x_i, x_j) = \exp\left(- \|x_i - x_j\|^2 \Big/ \sigma^2\right)\,. \tag{8}$$

Where $\|x_i - x_j\|$ is a norm of $x_i - x_j$ vector expressing the distance between x_i and x_j, and σ is a normalizing parameter, expressing the width of function surrounding central point.

2.2. Improved PSO algorithm

POS originates from the predatory behaviors of bird flock [13], that's, to find the best scheme through the information sharing and coordination behaviors between individuals of bird flock, which is regarded as an intelligent optimization method. Suppose that m particles exist in a d-dimensional search space, and each particle is a d-dimensional column vector, which can be expressed as $X_i = (x_{i1}, x_{i2}, \cdots, x_{id})$, $i = 1, 2, \cdots, m$, and the corresponding flying speed is $V_i = (v_{i1}, v_{i2}, \cdots, v_{id})$, $i =$

$1, 2, \cdots, m$, corresponding current optimal position is $P_i = (p_{i1}, p_{i2}, \cdots, p_{id}), i = 1, 2, \cdots, m$, and the current optimal position of population is $P_g = (p_{g1}, p_{g2}, \cdots, p_{gd})$. PSO algorithm is used to constantly update the position and speed of particles, and the speed update formula is as shown in equation (9):

$$v_{id} = \omega v_{id} + c_1 r_1 (p_{id} - x_{id}) + c_2 r_2 (p_{gd} - x_{id}) \,. \tag{9}$$

Where ω is an inertia weight coefficient, c_1, c_2 are acceleration constants and nonnegative numbers, determining the speed level of particles in converging to local minimum, r_1, r_2 are random numbers between 0 and 1. The position update formula is as shown in equation (10):

$$x_{id} = x_{id} + v_{id} \,. \tag{10}$$

However, PSO algorithm has the shortcoming of premature convergence, for which the method of controlling population diversity [13] is used to improve algorithm performance, and the implementation methods include initial population selection and premature convergence judgment [14] the two aspects.

The initial population of PSO algorithm is expressed by random matrix, where the particle position and speed parameter are within a certain interval. Under ideal conditions, the initial position of particles should be spread all over the space to increase the searching probability of global optimal solution. However, due to the limited number of particles, particles can not be guaranteed evenly spread over the solution space in general, which might increase the probability of falling into local optimal. Therefore, average particle distance $D(t)$ is introduced representing the dispersion degree of population distribution of particle swarm, the bigger the value is, the more disperse will the particle swarm become; otherwise, the more concentrated will be particle swarm become. Suppose dispersion threshold value β limits the value of $D(t)$, get particle swarm spread over the whole search interval as much as possible, and suppose the average particle distance is $d(t)$, then it can be got that:

$$d(t) = \frac{1}{mL} \sum_{i=1}^{m} \sqrt{\sum_{d=1}^{n} (p_{id} - \bar{p}_d)^2} \,. \tag{11}$$

Where, L is the maximum length value of opposite value in the search interval, m is the individual number of particle swarm, and n is the dimension of solution space, p_{id} is the coordinate vector of ith particle, and $\bar{p}_d$ is the average of all particles' coordinate vectors.

As the particle position in PSO constantly changes, the difference between particles are reduced gradually, and the coordinate of particles determine the corresponding fitness, therefore, the current update situation of particle swarm can be judged by combining with the fitness situation of particle swarm. Suppose the average fitness of particle swarm is $\bar{f}$ and the fitness of ith particle is f_i, it can be got that the fitness variance δ^2 of particle swarm is:

$$\delta^2 = \sum_{i=1}^{m} \left(\frac{f_i - \bar{f}}{f} \right)^2 . \tag{12}$$

Where, m is the individual number of particle swarm, f is the normalization scaling factor, with the function of limiting the size of fitness variance, and its calculation formula is:

$$f = \begin{cases} \max\left|f_i - \bar{f}\right| & \max\left|f_i - \bar{f}\right| > 1 \\ 1 & others \end{cases} \tag{13}$$

$$f = \begin{cases} \max\left|f_i - \bar{f}\right| & \max\left|f_i - \bar{f}\right| > 1 \\ 1 & others \end{cases} \tag{14}$$

Similar to variance concept, fitness variance represents the aggregation degree of particle swarm position, and the smaller the δ^2 value is, the more concentrated the particle position will be; otherwise, the more disperse particles will be. In the PSO algorithm, as the particles constantly update, δ^2 value will be reduced gradually, and when δ^2 value is lower than the preset threshold value, the search stage of decision algorithm enters into later stage, and since then, particle swarm easily falls into local optimal and presents the characteristic of premature convergence.

The two parameters $d(t)$ and δ^2 determined to measure particle swarm diversity respectively represent the dispersion degree between particles and the overall distribution situation of particle swarm, and during the update process of particle swarm, if particle swarm converges at several local minimal points and the corresponding fitness difference is relatively small, then in the two parameters, $d(t)$ is larger than δ^2. These two parameters can be used to judge whether particle swarm can be premature and then stop updating at search stage.

2.3. IPSO-LSSVM principle and procedure

In case that the classification model of LSSVM model is radial basis function, the parameters needed to be selected at initial stage are normalization parameter c and standardization parameter σ. If c is too big, the training error will be reduced and generalization ability of learning machine will be poorer; and if c is too small, the training error will be increased and generalization ability will be better; however, if σ is too big, lack of training and difficult localization will be caused, and if σ is too small, localization will be easier, but resulting in training risks. Therefore, an improved PSO algorithm is adopted to optimize LSSVM with the basic thought that equalize c and σ as two particles, and constantly update their speeds and positions, and use objective function to determine the corresponding fitness results, then through comparison, the global optimum results of c and σ are got. During the updating process of particle swarm, control the diversity of particle swarm based on the fitness variance δ^2 and average particle distance $d(t)$ to improve the overall performance of algorithm, so as to improve the performance of network intrusion detection model based on IPSO-LSSVM.

The algorithm steps for the network intrusion detection model based on IPSO-LSSVM are:

(1) Pretreat the historical network information data collected, map the symbol attribute value as numerical value and unify them into intervals within the interval of $[0.0, 1.0]$, and then establish the test samples set and training sample set;

(2) Initialize population size to generate multiple groups of particles, and generate random initialization speed and initialization position; according to equation (11), calculate the average particle distance $d(t)$ of each group of particles and select the group of particles with biggest $d(t)$;

(3) Get the fitness of each particle of the group, and through the comparison, update the optimum position P_i of individual and the global optimum position P_g in the population;

(4) Use equations (9) and (10) to update the speed and position of each particle;

(5) Use equation (12) to calculate the fitness variance δ^2 of particle and remove the position of premature convergence at later iteration period;

(6) Check to see whether the number of iterations is met, and if the maximum iteration number is met, the algorithm will be ended, output the normalization parameter c and standardization parameter σ after optimization; otherwise, shift into (3) to optimize individual position and speed;

(7) Use the optimized normalization parameter c and standardization parameter σ to construct LSSVM model and carry out model test.

At present, the most representative 4 kinds of attack states in the network include Probe attack, R2L attack and Dos attack. And in the IPSO-LSSVM detection model in this Paper, four kinds of classifiers are constructed to determine the five states of current network, namely four attack states and a normal state; IPSO-LSSVM, IPSO-LSSVM2, and IPSO-LSSVM3, IPSO-LSSVM4 classifiers are used successively to determine network state difference, with the use rules including: (1) if the output results of four classifiers are all 1, it will indicate the network states are respectively normal, Probe attack, R2L attack and U2R attack; (2) if the output result of front classifier is 1, then the next classifier judgment will not happen; and if the output results of four classifiers are all 0, it will be determined as Dos attack. In this way, IPSO-LSSVM detection model completes the detection and determination of network intrusion types.

3. Simulation experiment

3.1. Data selection and pretreatment

In order to test the performance of the IPSO-LSSVM model, the DARPA intrusion detection evaluation data from the MIT Lincoln Laboratory was used as the experimental data [8]. The integrated data of these data came from the simulated real network environment, where 10,000 normal data were randomly selected, and besides, there were 70 Probe attack samples, 70 R2L attack samples, 90 U2R attack samples and 200 Dos attack samples. Preprocess the data set, and map the symbol attribute value to numerical value firstly to complete data depiction and then unify it into the interval of $[0.0, 1.0]$, followed by random sampling on data to form training data subset and test data subset.

3.2. Parameter setting and index selection

After combining with the selection rules of normalization parameters, standardization parameters and inertia weight in literatures [5], [7]-[13], the value range of normalization parameter in the simulation verification of the Paper is $C \in [0, 100]$, that of standardization parameter is $\sigma \in [0, 10]$ and that of inertia weight is $\omega \in [0.4, 0.9]$. In the test, it is found that when the iterations of particle swarm algorithm exceed 100, performance tends to be stable, therefore, the maximum iterations number for particle swarm is 100; taken the number of particle swarm individuals as 20, and through the simulation analysis, it is found that when both c_1, c_2 are taken as 2, optimization effect is found best, therefore, c_1, c_2 values are taken as 2.

For quantitative comparison, select training time T_t, detection time T_c, detection accuracy P and false alarm rate Q, of which T_t and T_c are respectively the time of constructing model and the time of intrusion detection by using training sample set, and the calculation method of P is:

$$P = n_i / N_i \,. \tag{15}$$

Where, n_i is the number of intrusion samples detected, N_i is the total number of intrusion samples. And the calculation method of Q is:

$$Q = n_{iw} / N \,. \tag{16}$$

Where, n_{iw} is the number of normal samples detected as intrusion samples, and N is the number of normal samples.

3.3. Comparative analysis of experimental results

Comparing the detection accuracy and false alarm rate of IPSO-LSSVM model with the experimental results by using PSO-LSSVM, LSSVM and SVM models respectively, it is found that the initial parameter of particle swarm in PSO-LSSVM model is generated randomly, with its detection accuracy as shown in Fig. 1.

In Fig. 1 and Fig. 2, the X-coordinates are IPSO-LSSVM, PSO-LSSVM, LSSVM and SVM four kinds of network intrusion detection algorithms, and Y-coordinates are the detection accuracy and false alarm rate of each intrusion detection algorithm corresponding to Probe attack, R2L attack, U2R attack and Dos attack. Seen from Fig. 1: for various network attack ways, the detection accuracies of SVM, LSSVM, PSO-LSSVM and IPSO-LSSVM are improved in turn; and seen from Fig. 2: for various attack models, the false alarm rates of SVM, LSSVM, PSO-LSSVM and IPSO-LSSVM are decreased in turn; and the detection accuracy and false alarm rate of IPSO-LSSVM model are the optimal among the four models. Comparing the experimental results of various detection models: aiming at Dos attack and

U2R attack, the detection accuracy of LSSVM is improved by 0.2% than that of SVM, showing relatively small improvement range of detection accuracy; aiming at Probe attack and R2L attack, the detection accuracies of LSSVM are respectively improved by 2.4% and 1.7% than that of SVM, showing relatively big improvement amplitude, which is related to attack principle; compared with the detection accuracy of LSSVM, the detection accuracy of IPSO-LSSVM is improved by 4.3%-9.6%, indicating that it is feasible to use IPSO to seek global optimum value so as to realize the optimization of LSSVM model, and the detection accuracy of LSSVM after optimization is improved; compared with the detection accuracy of PSO-LSSVM, the detection accuracy of IPSO-LSSVM is improved by 2.1%-5.5%, indicating a further improvement is obtained in PSO' s optimization on LSSVM which is realized by controlling the diversity of particle swarm based on the fitness variance δ^2 and average particle distance $d(t)$ in particle swarm updating process. To sum up: using LSSVM to replace SVM and using IPSO to optimize LSSVM is feasible; due that premature convergence is avoided and the initial particle distance of PSO is improved after optimization, the use of LSSVM algorithm has improved the detection accuracy and false alarm rate of the model.

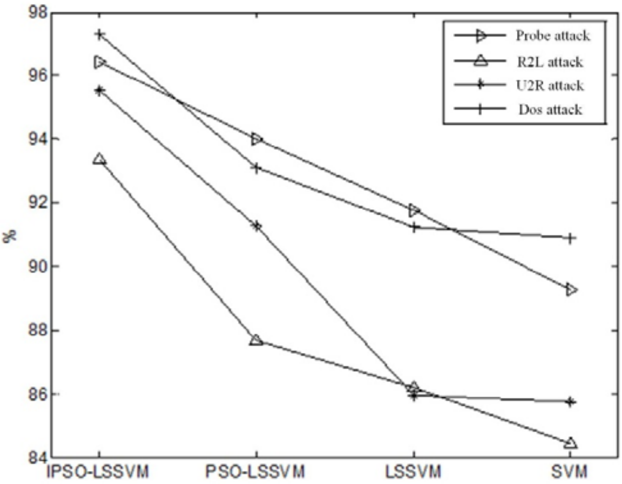

Fig. 1. Comparison of detection accuracy of four models

Compare the training time and test time of IPSO-LSSVM model and the experimental results of IPSO-SVM model, and the training time and test time results are as shown in Table 1.

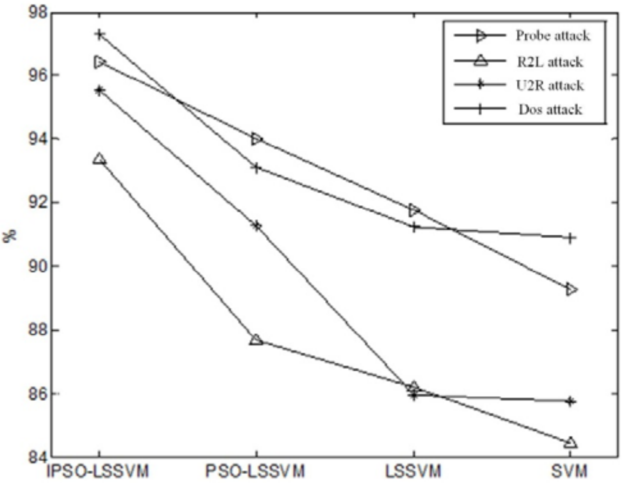

Fig. 2. Comparison of false alarm rate of four models

Table 1. Comparison of run time of two models (unit: s)

Intrusion types	IPSO-LSSVM		IPSO-SVM	
	Training time	Test time	Training time	Test time
Probe attack	3.74	0.81	6.07	0.98
R2L attack	3.52	1.21	5.81	1.16
U2R attack	5.29	1.69	8.17	1.75
DoS attack	8.41	1.79	18.21	2.04

Seen from Table 1, for the same kind of network attack, both the training time and test time of IPSO-LSSVM model are lower than that of IPSO-SVM, indicating that using LSSVM to replace SVM and using a new quadratic loss function to simplify the quadratic programming function in SVM into solving linear equation can effectively reduce the computation complexity and speed up computing speed, so that the run efficiency is improved at the same time of improving the detection accuracy of algorithm.

Compare the detection performances of intrusion detection algorithm of IPSO-LSSVM model with that of the intrusion detection algorithm of LSSVM model (ACO-LSSVM) based on ant colony optimization, and the results are shown in Table 2.

Table 2. Comparison of detection performance of two models

Intrusion types	Detection accuracy(%)		False alarm rate(%)	
	IPSO-LSSVM	ACO-LSSVM	IPSO-LSSVM	ACO-LSSVM
Probeattack	96.43	95.39	3.57	4.61
R2L attack	93.37	87.43	6.63	12.57
U2R attack	95.51	92.27	4.49	7.73
DoS attack	97.29	94.28	2.79	5.72

Seen from Table 2, the detection accuracy of the IPSO-LSSVM model is higher than that of the ACO-LSSVM model, and its false alarm rate is lower than that of the ACO-LSSVM model, and its detection performance is better than that of the ACO-LSSVM model, indicating that using IPSO can make particles seek the optimal parameter pair in the global scope, thus improving the detection performance of LSSVM model.

4. Conclusion

In order to improve the performance and efficiency of network intrusion detection, LSSVM model is used and the initial parameters of LSSVM model are optimized by using the improved PSO algorithm. And the results of verification test show that using LSSVM model can achieve higher efficiency than SVM model, and using IPSO-LSSVM model can achieve higher detection accuracy than PSO-LSSVM, LSSVM, SVM and ACO-LSSVM models; besides, the feasibility of optimization algorithm is verified, which provides a new research thought for network intrusion detection.

References

[1] M. R. Raman, N. Somu, K. Kirthivasan, et al.: *A Hypergraph and Arithmetic Residue-based Probabilistic Neural Network for classification in Intrusion Detection Systems.*[J]. Neural Networks, 2017 (2017), 92.

[2] M. R. G. Raman, K. Kirthivasan, V. S. S. Sriram: *Development of Rough Set – Hypergraph Technique for Key Feature Identification in Intrusion Detection Systems*[J]. Computers & Electrical Engineering. (2017).

[3] A. Guzzo, A. Pugliese, A. Rullo, et al.: *Intrusion Detection with Hypergraph-Based Attack Models*[C]// Revised Selected Papers of the Third International Workshop on Graph Structures for Knowledge Representation and Reasoning. Springer-Verlag New York, Inc. *2013* (2013), 58–73.

[4] M. R. G. Raman, K. Kannan, S. K. Pal, et al.: *Rough Set-hypergraph-based Feature Selection Approach for Intrusion Detection Systems*[J]. Defence Science Journal, *66* (2016), No. 6, 612.

[5] J. Chen, Y. Lin, G. Lin, et al.: *Attribute reduction of covering decision systems by hypergraph model*[J]. Knowledge-Based Systems, *2016* (2016), 118.

[6] J. Silva, R. Willett: *Hypergraph-based anomaly detection of high-dimensional co-occurrences*[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, *31* (2009), No. 3, 563–9.

[7] J. Che, W. Lin, Y. Yu, et al.: *Optimized Hypergraph Clustering-based Network Security Log Mining*[J]. Physics Procedia, *24* (2012), No. 1, 762–768.

[8] B. Luo, J. Xia: (2014) *A novel intrusion detection system based on feature generation with visualization strategy*[J]. Expert Systems with Applications, 41(9):4139-4147.

[9] Rital S., Cherifi H.: *Similarity hypergraph representation for impulsive noise reduction*[C]// Video/image Processing and Multimedia Communications, 2003. Eurasip Conference Focused on. IEEE, *2* (2003), 539–544.

[10] R. Qian, W. Zhang, B. Yang: *Community Detection in Scale-Free Networks Based on Hypergraph Model*[C]// Pacific Asia Conference on Intelligence and Security Informatics. Springer-Verlag, *2007* (2007), 226–231.

[11] X. Zheng, Y. Luo, L. Sun, et al.: *A novel social network hybrid recommender*

system based on hypergraph topologic structure[J]. World Wide Web-internet & Web Information Systems, *2017* (2017), 1–29.

[12] D. Tisza, A. Oláh, J. Levendovszky: *Novel Algorithms for Quadratic Programming by Using Hypergraph Representations*[J]. Wireless Personal Communications, *77* (2014), No. 3, 2305–2339.

[13] Y. Z. Li, D. Wu, J. D. Ren, et al.: *An Improved Outlier Detection Method in High-dimension Based on Weighted Hypergraph*[C]// International Symposium on Electronic Commerce and Security. IEEE, *2009* (2009), 159–163.

[14] K. Pahlavan, P. Krishnamurthy, Y. Geng: *Localization challenges for the emergence of the smart workd.* IEEE Access, *3* (2015), No. 1, 3058–3067

[15] Z. Lv, A. Tek, F. Da Silva, C. Empereur-Mot, M. Chavent, & M. Baaden: *Game on, science-how video game technology may help biologists tackle visualization challenges.* PloS one, *8* (2013). No. 3, e57990.

[16] Z. Lv, A. Halawani, S. Feng, H. Li, & S. U. Réhman: *Multimodal hand and foot gesture interaction for handheld devices.* ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), *11* (2014), No. 1s, 10.

[17] W. S. Pan, S. Z. Chen, Z. Y. Feng: *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences, *7* (2013), No. 2, 675–681.

[18] Y. Y. Zhang, J. W. Chan, A. Moretti, And K. E. Uhrich: *Designing Polymers with Sugar-based Advantages for Bioactive Delivery Applications*, Journal of Controlled Release, *219* (2015), 355–368.

Received May 7, 2017