*i*SPm$_A$: A Novel IOT Security Event Perception Model based on Autonomic Computing

Ruijuan Zheng, Tenghao Li, Mingchuan Zhang, Qingtao Wu, Zhengchao Ma, Wangyang Wei and Chunlei Yang

Information Engineering College, Henan University of Science and Technology, Luoyang 471003, China

Keywords: Internet of Things, Autonomic Computing, Principal Component Analysis, Feature Extraction

Abstract. According to the high dimension of security event feature and the difficulty of security event autonomic perception in the age of big data, a novel Internet of Things (IOT , for short) security event perception model based on Autonomic Computing and Principal Component Analysis (PCA, for short) is proposed, including element extraction, element understanding and event prediction. In which, to improve the real-time performance of element understanding, PCA is adopted to map the initial high dimensional feature to a set of new unrelated synthesized feature, and back propagation neural network (BP neural network, for short) is used to fuse the synthesized feature after reduction. The experimental result show that, feature reduction by PCA can greatly reduce the input dimension of fusion engine, efficiently cuts down the learning time of BP neural network, and improves the accuracy of event perception.

I. Introduction

As a kind of large scale network, the high hybrid, high heterogeneity and high uncertainty [1] of IOT [2] bring a big challenge to the system security. Together with the rapid development of IOT application, the speedy perception of system security event has become a major subject. At present, many researches on network security technology are becoming increasingly mature but there are still many deficiencies. Security event perception of IOT is a kind of network security state monitoring and detection technology to network administrator, which will effectively improve the network system emergency response capacity, reduce the damage of threats posed from cyber attacks and find potential malicious intrusion behavior. However, because of the growth of IOT scale, the security event feature shows the characteristics of huge volume, various types, low density of value, and velocity processing.

For the above characteristics, the IOT event perception becomes exceptionally complex. In the complex and fast-moving security environment, by giving full play to the advantages of IOT and coupling the Autonomic Computing thinking [3] and security policy, autonomic perception of IOT security event endows the IOT system with the ability to self aggregate, understand and show large scale of security elements that may Influence the security of network system. And by self judgment of autonomic security, IOT system will implement the self perception of system security event under complex network morphology [4], reducing the frequency and intensity of human intervention.

IOT is a kind of heterogeneous network, and its security event perception must be integrated from a large number of heterogeneous distributed sensors and security equipment data. Accurate perception of security events is the base to ensure the security of IOT application, as a result, the development of security event perception of IOT has put forward new requirement and challenge to the machine learning from research direction, evaluation index and the key technology. Firstly, the number of training instances is great and a large number of security data sets are converged every day. Secondly, including sensors, more and more security equipments continuously record observation data (which can be used as the training data set), and the data set will easily reach to hundreds of TB scale, with the feature of high dimension. It’s not feasible to input the data with high dimensionality to fusion engine, and the feature extraction(FE, for short) is inevitable. To implement FE, Lippmann proposed a keyword selection algorithm [5], and Zhang *et al.* realized a FE algorithm adopting rough set theory [6]. Kush used alternating minimization scheme to realize the synchronization of dimension reduction and learning [7], and Zhang adopted a multi metric learning algorithm to learn the optimal set of intraspecies and interspecies metrics, fusing the data [8] collected from multiple sensors. These algorithms give some inspiration to the network security event perception, but they are relatively complex, and will lose some information included in some feature.

As the core parts of data engine, machine learning and data fusion are the foundation to realize accurate perception of events, where, machine learning adopts the static or dynamic feature to train [9]. Bass [10] is the first one applying data fusion technology to network security, who proposes an intrusion detection system based on data fusion of multi sensor, implementing the detection, assessment to intrusion threat. Verma chose the feature selection based on correlation coefficient as the input of feed forward artificial neural network, proposed an online security assessment and contingency analysis measure to power systems using the supervised learning method [11]. To online predict the users' satisfaction degree on network application, adopting learning method based on support vector machine, Joumblatt *et al.* built a prediction model [12] , while, to different online users the method needed multiple training and was not universal.

Based on current research status, according to the IOT heterogeneity, complexity and fuzziness and focusing on the self perception problems of network security event, combined with the characteristics of IOT system security event index and based on the principle of Autonomic Computing thought, this paper proposes a network security event perception model based on Autonomic Computing. Here, PCA [13] is adopted to extract the security event attributes, and the comprehensive attributes are input into fusion engine.

II. MODEL FRAMEWORK

On the current research on IOT security event perception, there is not a general model. In view of this, this paper proposed an IOT Security Event Perception model based on Autonomic Computing (*i*SPm$_A$) inspired by the principle of network security situation awareness model. And the module of *i*SPm$_A$ is showed in Fig. 1.

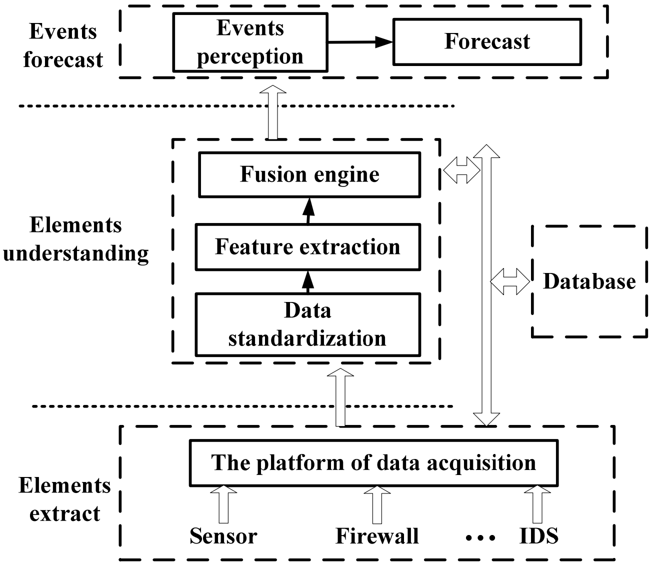

Fig. 1 The module of *i*SPm$_A$

*i*SPm$_A$ can be divided into three layers.

The first layer is for the perception element extraction (PE, for short). PE is a prerequisite for the security event perception, and the implementation of IOT security event perception system requires the fusion of multi-source heterogeneous data, as a result, PE needs to collect data from security equipments deployed in all layers of IOT and complete the data preprocessing.

The second layer is for the element understanding (EU, for short). And the EU layer is the core part for implementing the security event perception. In which, feature extraction is to find the attribute characteristics that can be easily used to distinguish attack behavior from the huge amount of network data attribute by using a certain method; by the classification of inherently linked data, machine learning is to determine the data membership of the attack behavior and realize self perception of security events.

The third layer is for event forecast (EF, for short). At the same time of implementing EU, some development forecast of future security events must be achieved according to current element characteristics.

III. Proposed Routing Strategy and Algorithm

The development of IOT security event perception has put forward new requirement and challenge to the machine learning from the research direction, evaluation index and key technology. By introducing the thought of data fusion and machine learning, the *i*SPm$_A$ system will implement the autonomic perception and accurate detection to security incidents. The *i*SPm$_A$ can be reflected through the following key technique.

A. Feature Extraction based on PCA

Under the assumption of reducing as far as possible the feature space dimension, feature extraction is proposed to avoid the fact that the fusion of large amount of data may cause the system detection rate cannot meet the real-time detection requirement of high speed networks, without reducing the accuracy of classification. Every day, IOT converges a large number of high dimension security event perception data, in which, "dimension disaster" is a big barrier to the application of IOT security event perception method to the actual system, because the input vectors with high dimensionality will bring a huge amount of computation which will make training methods to the feature set lose real time property. The purpose of feature extraction is to identify the features that are important to fusion and delete features that have little or no effect to the fusion result. Its essence is to find a subset of input features, and this subset is the feature set that affects the fusion results greatly.

PCA tries to recombine a host of initial variables to a new group of independent integrated variables, to replace the original ones, namely, to achieve a few representative variables from multiple variables, which can represent most information of original ones, and are uncorrelated [13-15]. Learning to the reduced feature can greatly reduce the learning time and resource utilization, and improve the response time of the system, in condition of ensuring the learning accuracy. The specific process of PCA is shown in Fig. 2.

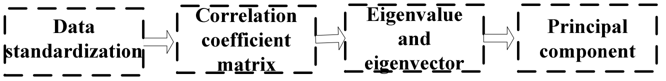

Fig. 2 Feature extraction process of PCA

The extracted attribute features are sequentially combined as the feature set *X*, and the data information of *P*-dimensional attribute feature in security event *X* is obtained.

$$X = (x_1, x_2, \cdots, x_p) \qquad (1)$$

P feature variables are $x_1, x_2, \cdots, x_p$,and the data array of *n* samples is:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \left(x_1, x_2, \cdots x_p\right),$$

$$x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}, j = 1, 2, \cdots p \qquad (2)$$

Data standardization. The calculation process is showed in formula (3).

$$x_{ij}^{*}=\frac{x_{ij}-\overline{x}_j}{\sqrt{\operatorname{var}(x_j)}}(i=1,2,\cdots,n;j=1,2,\cdots,p)$$

$$\overline{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij} \qquad (3)$$

$$\operatorname{var}(x_j)=\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\overline{x}_j)^2$$

Calculation to correlation coefficient matrix of samples. That is

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \qquad (4)$$

Time Assume that the original data after standardization is still expressed as *X*, then the correlation coefficient of standardized data is showed in formula (5).

$$r_{ij}=\frac{1}{n-1}\sum_{t=1}^{n}x_{ti}x_{tj}(i,j=1,2,\cdots,p) \qquad (5)$$

3) Calculation to the eigenvalue $\lambda_1,\lambda_2\cdots\lambda_p$ of correlation coefficient matrix *R* and responding eigenvector $a_i=\left(a_{i1},a_{i2},\cdots a_{ip}\right),i=1,2\cdots p$.

B. Selection to Principal Component

By PCA calculation, *p* principal components will be achieved, however, because the variance of each principal component is decreasing, the amount of information contained is declining, so actual analysis generally doesn’t select *p* principal components, but select the front *k* ones according to the cumulative contribution rate of each principal component. Here, contribution rate is the proportion of the variance of a principal component in the total variance, is actually the proportion of an eigenvalue in the total value of all eigenvalue. That is formula (6).

$$\text{Contribution Rate}=\frac{\lambda_i}{\sum_{i=1}^{p}\lambda_i} \qquad (6)$$

Larger contribution indicates that the information of initial variable included in a principal component is stronger. Selection of the principle components number *k* mainly depends on the cumulative contribution rate of principle components. Generally speaking, contribution rate of the selected ones is more than 80% or the eigenvalue of selected principal component is bigger than 1, so as to ensure that the selected integrated variables can include most information of initial variables.

The resulting *k* feature vector consists of the following linear transformation, that is

$$\begin{cases} F_1=a_{11}x_1+a_{12}x_2+\cdots+a_{1p}x_p \\ F_2=a_{21}x_1+a_{22}x_2+\cdots+a_{2p}x_p \\ \cdots \\ F_k=a_{k1}x_1+a_{k2}x_2+\cdots+a_{kp}x_p \end{cases} \qquad (7)$$

It can be shortened as formula (8).

$$F_j=\alpha_{j1}x_1+\alpha_{j2}x_2+\cdots+\alpha_{jp}x_p(j=1,2,\cdots,\mathrm{k}) \qquad (8)$$

Then, F_1 is the first principal component, F_2 is the second principal component, by analogy, there will be *k* principal components, and a_{ij} is the principal component coefficient. The model can be expressed as a matrix $F=AX$, which is a matrix of principal component coefficient. The training data set can be substituted to the principal component matrix to obtain the reduced data set.

C. Fusion engine of Back Propagation Neural Network

From the angle of view of fusion method, at present the Bayesian theory, artificial neural network and support vector machine is widely used in data fusion. In this *i*SPm$_A$ model, BP neural network is selected as the fusion engine. BP neural network [16] has a lot of advantages, such as nonlinear mapping [17], self-learning, adaptive ability, easy to realize the parallel computation, etc, so the nonlinear relationship mapping between the indexes and the evaluation results can better realized.

Neural network is a high parallel processing system composed of massive and widely connected neurons, in which, the output of neurons in a certain layer is as the input of neurons in lower layer, and the output layer in the neural network will output final value of the whole network. BP neural network is a widely used neural network model in many field applications. Fig. 3 shows a topological structure of BP neural network. BP neural network algorithm can be described as follows.

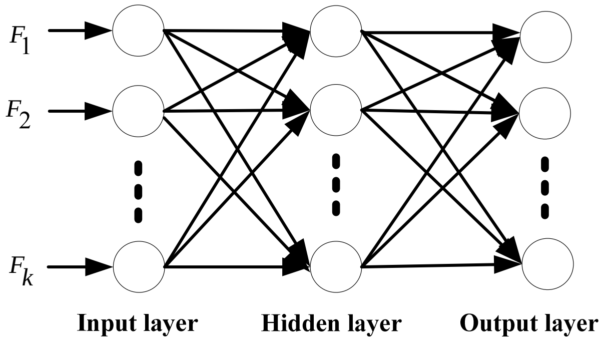

Fig. 3 Back propagation neural network topology structure

(1) Initialize the hidden layer and output layer weights.

(2) Input training vector X_n and the expected output vector O_k, in which, *n* and *k* is separately as the dimension of input vector and output vector. Then, iterate from step (3) to step (5).

(3) Calculate the output of neurons in hidden layer and output layer.

(4) Calculate the deviation between the neurons in hidden layer and output layer.

(5) Adjust the weights and thresholds of neurons.

(6) Judge the deviation accuracy, if it meets the requirements, the learning stops, or it will jump to step (2) to continue the iteration.

IV. Experimental Results and Analysis

A. Basic Configuration

The key to realize the *i*SPm$_A$ model includes element extraction and element understanding, so the core part of the experiment is to carry on the feature extraction and feature fusion of data sets. Experiments are carried out in the windows7 operating system, the programming environment uses MATLAB R2010b, and the data source selects KDDCUP.data_10_percent.gz intrusion detection data set. The training set and testing set of *i*SPm$_A$ are shown in table 1.

Tab. 1 Training and test sets

Type	Train set	Test set
Normal	981	1867
DOS	3084	7553
Probe	674	278
R2L	263	745
U2R	52	94

Before feather extraction, variance of each attribute in the training set will be calculated. According to the mathematical property of variance, if the variance of an attribute is 0, the attribute will be deleted directly. The attribute is mapped to a specific numerical value. Taking the second attribute protocol_type for example, TCP is set as 0.3, UDP is set as 0.6, and ICMP is set as 0.9. After the initial attribute selection, 35 attributes are obtained.

B. Performance Test

Firstly, feature extraction is executed according to the PCA method introduced in section 3.1, achieved eigenvalue of correlation coefficient matrix R is shown in Figure 4, and the cumulative contribution rate is shown in figure 5.

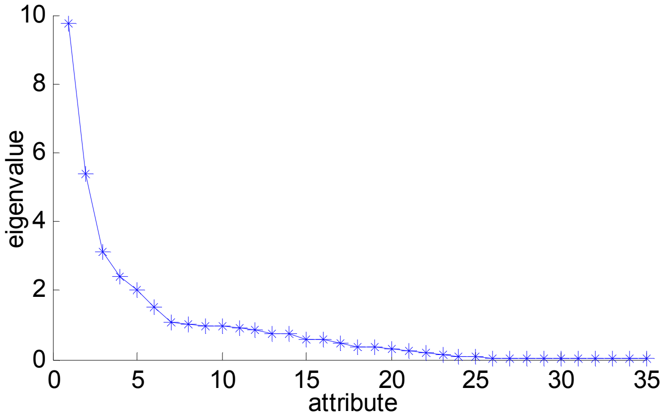

Fig. 4 Correlation coefficient matrix eigenvalues

As we can see from Figure 5, cumulative contribution rate of the front 11 eigenvalues reaches 83.44%, so the corresponding 11 principal components are selected. After feature extraction, the number of feature is reduced to 31.43% of that in the original properties, the amount of information is 83.44% of that in the original properties, and the feature dimension is greatly reduced.

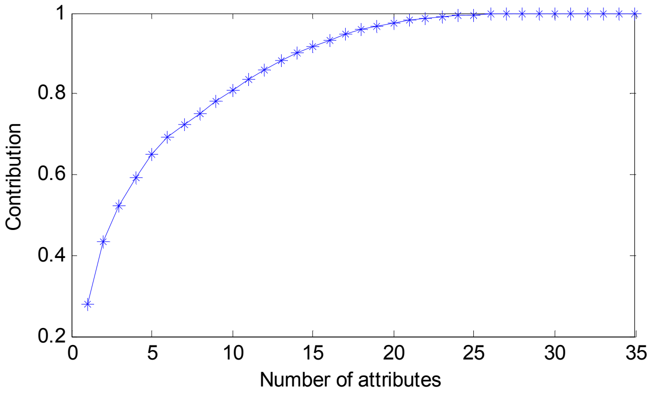

Fig. 5 Cumulative contribution rate

In order to facilitate the application of BP neural network to the *i*SPm$_A$, this paper creates a structure S={IL,HL,OL}. IL means the dimension of the input vector, HL represents the number of hidden neurons, and OL denotes the number of output neurons in neutral network. In order to facilitate the fusion, the input vector is formatted according to formula (9), so as to the component of each input vector is mapped to [0,1] interval.

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \qquad (9)$$

The final dimension of one data set identifies the connection record type, including Normal connection, DOS attack, Probe port scan attack, R2L attack and U2R attack. In the experiments, we map the five connection record types into five system security values (SV, for short) according to the corresponding harm degree. For normal connection records, the SV is 0.9; DOS attack is a means of attack that will make the system cannot provide normal services to other users by depleting system resources, so the SV is 0.1; Probe is refers to scan to the computer network or server, and the SV is 0.7; the R2L is an attack that remote user obtains the host access, and the SV is 0.5; the U2R is the attack that the local user gets administrator privilege, so the SV is 0.3.

The data set obtained by attribute extraction is called as BP neural network of principal component analysis (PCABPNN) subset, and the original data is called as BP neural network (BPNN) set. In order to ensure the accuracy of the experiments, this paper carries on 5 times of repeated experiments, after each experiment, the stored data in the program is emptied. Tab. 2 lists the mean training time of BPNN network before and after extraction in the 5 experiments. It can be seen from the table, the extracted attribute data reduces the network training time, thus the BP neural network has better real-time performance. Fig. 6 describes the average time comparison between full training set and test set before and after feature extraction in 5 training and testing, and the extracted attributes are better than all attributes in both training time and testing time.

Tab. 2 Time consumption

Attributes number	BPNN structure	Train set (s)
35	(35, 10, 1)	16.898
11	(11, 10, 1)	6.013

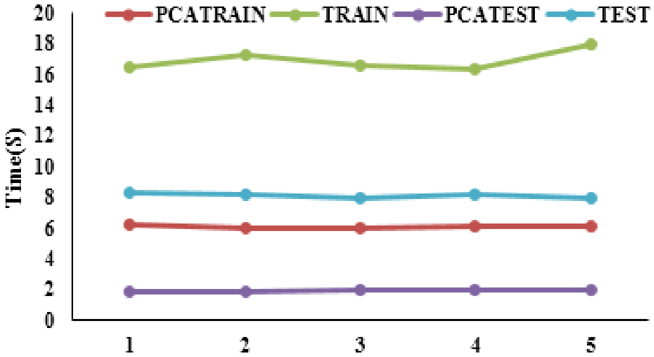

Fig. 6 Training and testing time

On the basis of feature extraction, to each kind of attack, the performance of event detection accuracy and detection time based on all attributes and attribute subset is compared. The experimental results are as shown in Fig. 7 and Fig. 8. It can be seen from Fig. 7 that attribute subset extracted by PCA method is better than all attributes in the detection accuracy. As you can see from Fig. 8, the integrated attribute extracted by PCA method is better than all attributes in the performance of detection time.

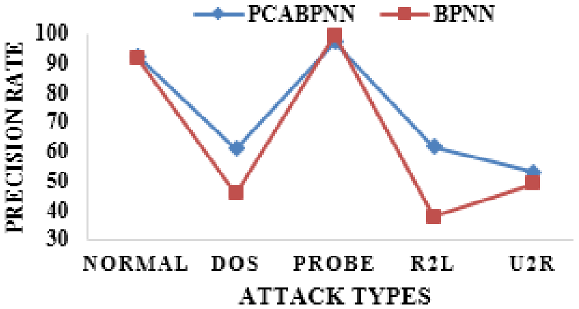

Fig. 7 Detection accuracy before and after feature extraction

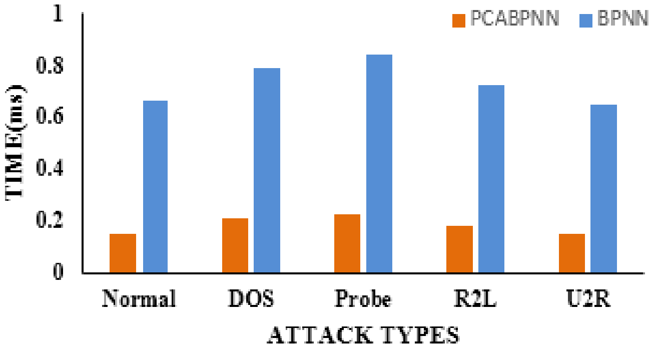

Fig. 8 Detection time before and after feature

From the experimental results, we can see that the proposed *i*SPm$_A$ effectively reduces the attribute dimension of perception information, test model established according to the extracted attributes has certain advantages over the model established using all attributes in the accurate detection rate and detection time.

V. CONCLUSION

According to the characteristics of perceived security events in IOT, such as high dimension, huge volume, various types, low density of value, and velocity processing, this paper introduced Autonomic Computing idea into IOT security event perception model, put forward an IOT security event perception model based on Autonomic Computing *i*SPm$_A$, providing the design thought, implementation process, related models and experiment results. Firstly, the model formats the perception data by perception element extraction module, then it implements feature reduction of formatted data using the method of PCA. After that, the reduced attribute is input into learning engine, and finally autonomous perception of IOT security incidents is realized, reducing the human judgment to security properties. The results of experiments show that feature extraction using PCA method reduces the network training time and improves the accuracy of event perception.

Because the original training data set and test data set contain a lot of noise data, the experiments just simulated the network security event perception, and did not join the relevant mechanisms to eliminate noise data, which will affect the experimental results to a certain extent. Removing noise data is the next research step to our group, in addition, prediction technology in *i*SPm$_A$ is the important research goal of the future research.

ACKNOWLEDGMENT

This work was supported by Project of Natural Science Foundation of China (No. 61142002, U1204614, 61003035), Natural Science Foundation of Henan Province Education Department (No.2010A520018), Key Project of Science and Technology Department of Henan Province (No.112102210187) and Research Funds for Doctor of Henan University of Science and Technology.

REFERENCES

[1] I. Strategy, P. Unit. “Itu internet reports 2005: The internet of things”, Geneva: International Telecommunication Union 2005.

[2] QB. Sun, J. Liu, S. Li, CX. Fan, JJ. Sun, “Internet of Things: Summarize on Concepts Architecture and Key Technology Problem”, Journal of Beijing University of Posts and Telecommunications, vol.33, no.7, pp.1-9, Dec. 2010.

[3] P. Horn, “Autonomic Computing: IBM’S Perspective On the State of Information Technology”, IBM Corporation, 2001.

[4] YL. Hu, YF. Sun, BC Yin, “Information sensing and interaction technology in Internet of Thing”, Chinese journal of computers, vol.35, no.6, pp.1147-1163, Aug. 2012.

[5] RP. Lippmann, RK. Cunningham, “Improving intrusion detection performance using keyword selection and neural networks”, Computer Networks, vol.34, no.4, pp.597-603, Oct. 2000.

[6] M. Zhang, JT. Yao, “A rough sets based approach to feature selection,” the Proc. of NAFIPS, pp.434-439, Banff, Alberta, Canada, Jun. 2004.

[7] KR. Varshney, AS. Willsky, “Linear dimensionality reduction for margin-based classification: High-dimensional data and sensor networks”, IEEE Transactions on Signal Processing, vol.59, no.6, pp.2496-2512, Jun. 2011

[8] YN. Zhang, HC. Zhang, NM. Nasrabadi, TS. Huang, “Multi-metric learning for multi-sensor fusion based classification”, Information Fusion, vol.14, no.4, pp.431-440, Oct. 2013.

[9] I. Santos, J. Devesa, F. Brezo, J. Nieves, PG. Bringas, “OPEM: A Static-Dynamic Approach for Machine-Learning-Based Malware Detection,” the Proc. of CISIS, pp.271-280, 2013.

[10]T. Bass, “Multisensor data fusion for next generation distributed intrusion detection systems”, IRIS National Symposium Draft, pp.1-6, Apr. 1999.

[11]K. Verma, KR. Niazi, “Supervised learning approach to online contingency screening and ranking in power systems”, International Journal of Electrical Power & Energy Systems, vol.38, no.1, pp.97-104, Jun. 2012.

[12] DZ. Joumblatt, J. Chandrashekar, B. Kevton, N. Taft, R. Teixeira, “Predicting user dissatisfaction with internet application performance at end-hosts,” the Proc. of INFOCOM (mini-conference), pp.1-5, Turin, Piemonte , Italy, Apr. 2013.

[13]YH. Kao, B. Van Roy, “Learning a factor model via regularized PCA”, Machine Learning, vol.91, no.3, pp.279-303, Jun. 2013.

[14]M. Humberstone, B. Wood, J. Henkel, JW. Hines, “Differentiating between expanded and fault conditions using principal component analysis”, Journal of Intelligent Manufacturing, vol.23, no.2, pp.179-188, Apr. 2013.

[15]VM. Janakiraman, XL. Nguyen, D. Assanis, “Nonlinear identification of a gasoline HCCI engine using neural networks coupled with principal component analysis”, Applied Soft Computing, vol.13, no.5, pp.2375-2389, May. 2013.

[16]YN. Zhang, DS. Guo, L. Zhan, “Common Nature of Learning Between Back-Propagation and Hopfield-Type Neural Networks for Generalized Matrix Inversion With Simplified Models”, IEEE Transactions on Neural Networks and Learning Systems, vol.24, no.4, pp.579-592, Apr. 2013

[17]J. Chen, YH. Qiu, “Multi-objective Optimization Design of Complex Production Processing Based on Genetic Algorithm and Neural Network”, Computer Science, vol.39, no.1, pp.215-218, Jan. 2012.