

Fault Detection and Identification of COSMED K4b² based on PCA and Neural Network

Jing Zhou, Steven Su, and Aihuang Guo

Abstract—COSMED K4b² is a portable electrical device designed to test pulmonary functions. It is ideal for many applications that need the measurement of the cardio-respiratory response either in the field or in the lab is capable with the capability to delivery real time data to a sink node or a PC base station with storing data in the memory at the same time. But the actual sensor outputs and data received may contain some errors, such as impulsive noise which can be related to sensors, low batteries, environment or disturbance in data acquisition process. These abnormal outputs might cause misinterpretations of exercise or living activities to persons being monitored. In our paper we propose an effective and feasible method to detect and identify errors in applications by principal component analysis (PCA) and a back propagation (BP) neural network.

Keywords—BP Neural Network, Exercising Testing, Fault Detection and Identification, Principal Component Analysis.

I. INTRODUCTION

THE expenditure of energy in sports or normal living activities is significant in analysis of physical activity epidemiology and sport performance. These parameters are traditionally tested through indirect calorimeter with a metabolic cart. However, the large dimension of most metabolic carts restricts these measurements to a laboratory environment. While many activities can be simulated in the laboratory (e.g. walking on a treadmill), other activities, involving in occupational and recreational activities, may not be duplicated in the laboratory. Using heavy equipments to collect expired air in the field for later analysis in the laboratory often interferes activities under investigation. Due to the improvement of miniaturized metabolic measurement systems consumption of oxygen can be measured outside the laboratory in natural circumstances. Recently lightweight, portable telemetric gas analysis systems are used to acquire measurements of daily activities and energy expenditure for further examination. One typical instrument is the COSMED K4b² metabolic measurement system [1]. It is designed to measure parameters of ventilatory, oxygen cost and carbon dioxide production with several sensors such as flow meter, oxygen sensor, carbon dioxide sensor, environment sensor and so on. This system is one of the latest portable devices for

cardiopulmonary gas exchange analysis base on true breath-by-breath without limitation.

One typical instrument is the COSMED K4b² metabolic measurement system. It is designed to measure parameters of ventilatory, oxygen cost and carbon dioxide production with several sensors such as flow meter, oxygen sensor, carbon dioxide sensor, environment sensor and so on. This system is one of the latest portable devices for cardiopulmonary gas exchange analysis base on true breath-by-breath without limitation.

Along with the widely usage of K4b², many research papers focused on the accuracy of k4b². But limited reports on the reliability of the COSMED K4b² system, especial in telemetry transmission, have been published. R Duffield assessed the validity and reliability of a COSMED K4b² portable telemetric gas analysis system by experiments and satisfactory reliability is formed in specific steady state and sustained maximal exercise [2]. Without doubt the reliability is also crucial for applications in metabolic testing or health monitoring. Apart for the errors caused by devices there are also a lot of fault due to transmission environment. The actual sensor outputs and data received by the central node or base station may contain some errors, such as impulsive noise which is related to sensors, low batteries, environment or errors in data transmission process. These abnormal outputs might cause misinterpretations of exercises or living activities to people being monitored. Therefore, an effective and feasible method to detect and identify errors in applications is significant and inevitable.

In this paper, we use principal component analysis (PCA) to detect fault and neural networks to identify the faulty sensors with reconstructing. The remaining sections of this paper are presented as follows. Section II introduces the function and applications of K4b². Section III explains PCA and its properties. In section IV, the principle of back propagation neural networks is outlined. Section V presents the process of fault detection and identification with interpretation simulation results. The section VI makes a conclusion.

II. OVERVIEW OF K4B² SYSTEM

A K4b² system is a COSMED portable medical instrument used for pulmonary function tests. It is utilized by physicians or trained people on a physician responsibility. It can be worn by the subject during activity, operates on battery power, and is capable of delivering real-time measurements in to a PC base station with saving data simultaneous. Because of its

Jing Zhou is with Faculty of Engineering and Information Technology, University of Technology, Sydney, 2007 Australia (phone: 02-9514-2376; e-mail: Jing.Zhou-2@student.uts.edu.au).

Steven Su is with Faculty of Engineering and Information Technology, University of Technology, Sydney, 2007 Australia (e-mail: Steven.Su@uts.edu.au).

Aihuang Guo is with School of Electronics and Information, Tongji University, 201804 China (e-mail: tjgah@ tongji.edu.cn).

convenience, it is accepted in many applications such as: sports medicine research in human performance, gait lab, occupational health, cardiology, cardiac rehabilitation, clinical nutrition. It can measure physiological response to exercise in the field without limitation. This telemetric gas analysis system is consisted of a soft, flexible face mask to sample expired air and a sensor unit to test ventilation (VE), oxygen (O₂) and carbon dioxide (CO₂) concentrations in the expired air.

There are three modes of data acquiring of K4b²:

(1) Holter Data: Recorder The unit stores breath by breath data. After the test, all data may be downloaded to the PC for analysis and presentation.

(2) Telemetry Transmission: Digital telemetry transmission is used to send breath by breath data to the PC located away from the subject (up to 800 meters). During transmissions K4b² stores all data in its memory to avoid data loss.

(3) Laboratory Station: Connect the RS-232 cable to a PC and turn the mobile K4b² into a conventional Metabolic Cart with same features of the best standalone labs.

The measured parameters of K4b² in breath by breath exercise testing are showed in Table I [3]:

TABLE I
 MEASURED PARAMETERS IN K4B²

Symbol	UM	Parameter
VO2	ml/min	Oxygen Uptake
VCO2	ml/min	Carbon Dioxide production
Vt	l	Tidal Volume
FetO2	%	End Tidal O2
FetCO2	%	End Tidal CO2
R	---	Respiratory Quotient
VE	l/min	Ventilation
HR	l/min	Heart Rate
Qt	l	Cardiac output
AT	---	Anaerobic Threshold
VE	l/min	Ventilation
SV	l/min	Stroke volume
RF	l/min	Respiratory Frequency
FeO2, FeCO2	%	Averaged expiratory concentration of O2 or CO2
VE/VO2	---	ventilatory equivalent for O2
VE/VCO2	---	ventilatory equivalent for CO2
VO2/HR	ml/beat	Oxygen pulse
VO2/Kg	ml/min/Kg	VO2 per Kg
Ti, Te, Ti/Ttot	sec	time breaths
Vd/Vt	---	Vd/Vt ratio
PaCo2	mmHg	arterial PCO2 (estimated)
P(a-et)CO2	mmHg	Delta PaCO2 – PetCO2

III. FAULTS DETECTION BY PCA

A. Principle of PCA

PCA is a vector space transformation often used to transform multivariable space into a subspace which preserves maximum variance of the original space in minimum number of dimensions [4]. The measured process variables are usually correlated to each other. PCA can be defined as a linear transformation of the original correlated data into a new set of uncorrelated data. Therefore PCA is a useful method to transform original process variables in a new set of uncorrelated variables that represent the trend of the process.

PCA is used for fault detection in many fields [5,6].

In normal condition, the PCA is established with a collected data matrix $X \in R^n$, where n is the number of samples and m is the number of variables. This matrix must be standardized to eliminate the effects of different units of variables. So the standard database is firstly normalized. Then construct the covariance matrix

$$R = \frac{1}{n-1} \bar{X}^T \bar{X} \quad (1)$$

And then perform the SVD decomposition on R:

$$R = U D_{\lambda} U^T \quad (2)$$

where $U_{m \times m}$ is a unitary matrix, and $D = \text{diag}(\lambda_{i=1,2,\dots,n})$ is a diagonal matrix. In equation 2, $U = [u_1, u_2, \dots, u_m]$ is a standard base of R_m and the database is described based upon U. The variances of in the every direction from the new coordinate satisfy $\lambda_1 > \lambda_2 > \dots > \lambda_n$, where $\lambda_{i=1,2,\dots,n}$ are the diagonal elements of D_{λ} . The subspace, which is formed with the first k ($k < n$) vectors without correlation, is called principal component subspace and the other subspace, which is formed with the n-k vectors $\tilde{S} = [u_{k+1}, u_{k+2}, \dots, u_n]$ are called residual subspace \tilde{S} . So the database X with m dimensions is replaced by the principal subspace \hat{S} with k dimension and the residual subspace \tilde{S} with n-k dimension.

The transformation matrix $P \in R^{n \times k}$ generated by chose k eigenvectors or columns of U related to k eigenvalues. Elements of T, called as scores, are calculated by Columns of matrix P.

$$T = XP \quad (3)$$

Scores are the values of the original measured variables that have been transformed into the reduced dimension space. Where $X = TPT + E$, $E = X - \hat{X}$ at last raw data space can be calculated as:

$$X = TPT + E \quad (4)$$

It is critical to select the number of principal components, because TPT represents principal elements of variability in the process and E stands for the variability related to process noise. The popular procedure for choosing components is Cumulative Percent Variance (CPV) approach. It is a method of the percent variance CPV (i) captured by the first k principal components:

$$CPV(i) = \frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^n \lambda_j} \quad (5)$$

B. Fault Detection with PCA

Fault detection can be realized in the subspace with fewer dimensions by Hotelling T^2 statistical variables. A PCA model is established according to historical data and multivariate control diagram set by Hotelling T^2 . It stands for the major variation in the data and is counted as the sum of squares of a new process data vector x :

$$T^2 = x^T P U_k^{-1} P \quad (6)$$

where U_k is a squared matrix formed by the first k rows and columns of U .

The data are regarded as right for a given significance level if: $T^2 \leq T_{\alpha}^2$.

$$T_{\alpha}^2 = k * (n-1) / n - k * f_{inv}(\alpha, k, n-k) \quad (7)$$

where $f_{inv}(\alpha, k, n-k)$ is the critical value of the Fisher-Snedecor distribution with $n, n-k$ freedom and α the level of significance. α usually uses values between 90% and 95%. T^2 comes from the first k principal components therefore it offers a detection for derivations in the latent variables which are most critical to the variance in the procedure.

IV. OVERVIEW OF BP NEURAL NETWORKS FIGURES AND TABLES

A. Principles of Neural Networks

An artificial neural network is regarded as a mathematical model to simulate some behaviors of biological nervous systems [7]. Neurons are connected together with weights so that they can deal with information collaboratively and store the information on these weights. The structure of neural network functions is denoted as Fig. 1: Each neuron receives a signal from the neurons in the previous layer, and every signal is multiplied by a separate weight value. The weighted inputs are added, and passed through a limiting function which scales the output to a fixed range of values. The output of the limiter is then broadcast to all of the neurons in the next layer.

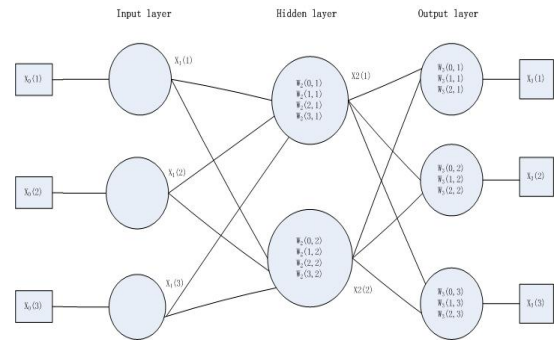


Fig. 1 Structure of a neural network

The inputs to the network are $x_0(1), x_0(2), x_0(3)$ up to $x_0(n)$. The outputs of the first layer, which are the inputs to the second layer, are $x_1(1), x_1(2), x_1(3)$ etc. Therefore $x_{j(i)}$ is the output from the i th neuron in the j th layer. The weights, which are called w , will denote the layer. For example: $w_{2(3,1)}$ indicates a weight in the second layer which is connected to $x_1(3)$, and which contributes to the output $x_2(1)$ via a non-linear function $f(\cdot)$ as:

$$x_2(1) = f(net_2(1)) \quad (8)$$

For a fully connected network, such as the one in figure 1, the value of $net_2(1)$ is:

$$net_2(1) = \sum_{i=0}^n w_2(i,1)x_j(i) \quad (9)$$

In order to train a multi-layered structure a learning rule is needed and one of the popular functions is:

$$y = \frac{1}{1 + e^{-x}} \quad (10)$$

B. Introduction of a BP Neural Network

In this paper, we use a BP neural network for fault reconstruction based on predicted values.

A BP neural network is one feed forward neural network, in which the neurons are arranged in layers, and each neuron can be connected only with the neurons in the next layer.

BP neural networks are trained by a supervised learning algorithm (e.g., an error back-propagation algorithm). An error back-propagation algorithm is an iterative procedure typically used to train a BP neural network. Specifically, the process modifies the weights in the network in an iterative fashion so that the resulting network fits the training data well. When expected outputs are known, a supervised learning algorithm can train the network more accurately and efficiently than an unsupervised learning algorithm can.

The entire network learning process includes two phases; the first stage is calculating from input layer to output layer. Output of all neurons can be calculated by training samples by initial structure and weight; the second stage is to modify the weights and threshold, and it start from output layer to input layer, and

weight of neurons connect to output can be adjusted according to errors of output, and also hidden layer weight can be modified too. The two stages are iterative process, repeat until convergence. All layers adjust weight through formula:

$$w_i(t+1) = w_i(t) + \eta \delta_i x_j^k \quad (11)$$

η is regarded as precision of network learning, which can be used as conditions of judgment of network finishing. δ_i is the value of error and can be defined as

$$\delta_i = y_{i(1)} - y_i \quad (12)$$

where y_i is the output value and d_i is the desired output.

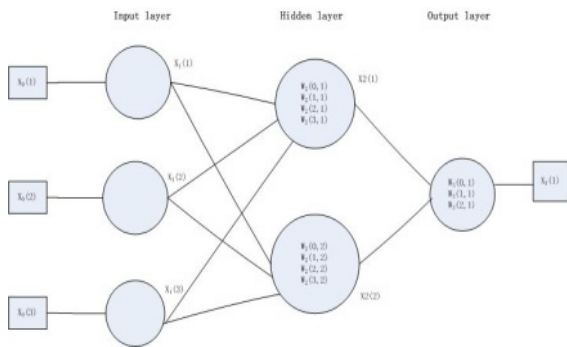


Fig. 2 A single output BP neural network

A single output BP neural network showed in Fig. 2 has been proved to be a useful model for prediction and forecasting. For example: Jiantao Liu used a BP artificial neural network to predict the flow stress of high-speed steel during hot deformation [8]. Yong Wang presents an accurate electricity load forecasting algorithm with back propagation neural networks [9].

V. PROPOSED METHOD AND SIMULATION RESULTS

A. Procedure of PCA Calculation and Fault Detection

From previous introduction we know PCA is effective to detect fault and BP neural network is capable to predict values for reconstructing. So we combine these two methods for fault detection and identification to input the reliability in applications of K4b². The whole process of fault detection and identification are outlined as following steps:

The first step: Build a PCA model based on acquired data and get the main eigenvalues for further analysis;

The second step: Calculate T² and compare with threshold to detect errors.

The third step: Train a BP neural network based on inputting data until they meet the defined criteria. When error is detected we use the trained BP neural network to reconstruct the possible value based on the measured value before the fault.

The fourth step: Compare the reconstructed data with real sampled data and identify the faults.

B. PCA Analysis and Fault Identification

In our experiment data are derived directly from the measured parameters in UTS laboratory for exercise monitoring. A K4b² system measured physical values such as VO₂, VCO₂, FetO₂, VE, HR and etc. Parts of the measured data are showed in Table II.

TABLE II
 MEASURED DATA BY K4B²

Rf	VT	VE	VO ₂	VCO ₂
b/min	l	l/min	ml/min	ml/min
8.298	1.2413387	10.301566	569.44046	483.4951
52.17391	1.3086587	68.277843	3115.2134	2650.346
32.78688	1.3433386	44.043889	1933.5084	1706.407
43.16546	1.3637386	58.866414	2438.8487	2181.350
33.51955	2.1144578	70.875681	2763.8671	2660.936

From the recorded data, we chose hundreds samples of 22 variables in the process and computing the feature matrix U for PCA analysis. The eigenvalues U_i and corresponding eigenvectors V of covariance matrix C of training samples X can be obtained, and further decreasingly ordered. The first k eigenvectors are packed to form the feature matrix base on equation (5). A calculation based on CPV reveals the first three components represent the high portion of data, which is clearly shown in Table III with the main eigenvalues and the proportion of each eigenvalue in total data.

TABLE III
 EIGENVALUES AND THE PROPORTION

Principal Components(PC)	Eigenvalue	Proportion
1	313138.1	0.976558
2	6740.755	0.021022
3	700.2923	0.002184
4	52.16869	0.000163
5	13.65424	4.26E-05

From this table we can clearly notice that the first three eigenvalues represent approximately 99% of the total data. Therefore, in our work we will take the first three principal components and use corresponding scores as parameters for fault detection.

Then we use this method to test sampled information and found that the 53rd parameter is abnormal compared with T square threshold in Fig.3. It means a fault happened at that time.

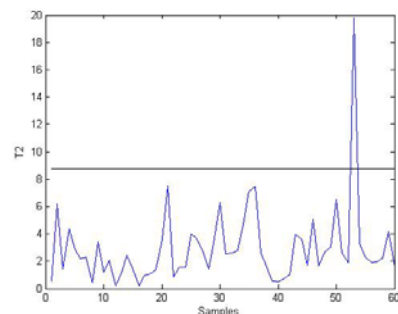


Fig. 3 T² Detection with transient error

C. Prediction and Fault Reconstruction

Then we build up a BP neural network for prediction with 10 input-layer neurons one hidden-layer neurons, and one output-layer neuron. The transfer function of each neuron is set to the sigmoid function as equation (6).

In order to make the BP neural network predict reliably, the BP neural network has to be trained properly. The following steps are enforced to train:

- (1) Provide 10 continuous normal samples to the neural network to generate the actual outputs.
- (2) Compute the error of the network by given corresponding output data. Then, propagate the error backward to the input layer. In this error back-propagation process, the weights on connections are changed to reduce the error. These steps will be repeated until the network error is small. A trained BP neural network is showed in Fig. 4.

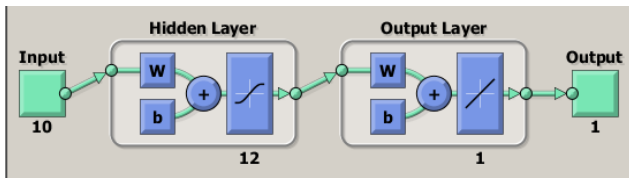


Fig. 4 Trained neural network

Afterwards we input 10 real time values, which comes from sampling 10 steps before the fault sampling point, to the trained network then get a predicted value.

We compare the predicted data with the fault samples in 22 values and it is clearly to see in Fig. 5 there is huge difference in the value of Vo_2 (The amount of oxygen consumed by the body each minute during a particular activity.) It is acquired by the oxygen sensor in the K4b² therefore it is possible this sensor is malfunction or the signal is interfered during transmission.

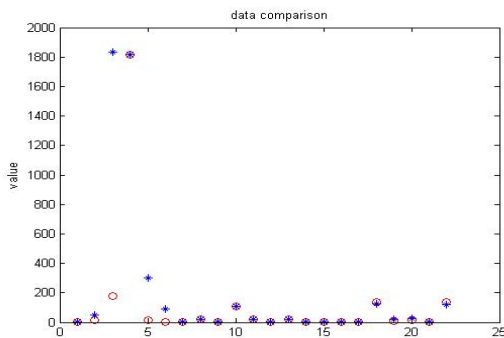


Fig. 5 Comparison measured data and predicted values in a transient error

Predicted variables are used to reconstruct the fault and are tested again with PCA. In Fig. 6 we can see that T^2 value dropped into control threshold and it proves the fault is excluded.

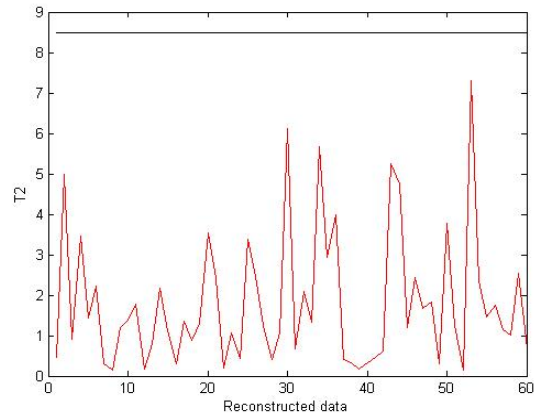


Fig. 6 T^2 Detection with reconstruction of a transient error

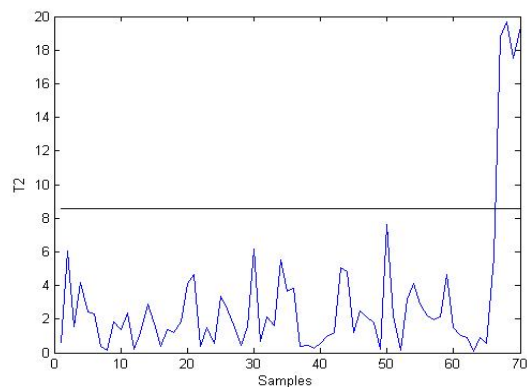


Fig. 7 T^2 Detection with continuous errors

Then we use this method again in different samples and detect continuous errors happened after the 67th sampling point as shown in Fig.7. Using the trained BP neural network to reconstruct errors and find that in Fig. 8 apart from the wrong value of value Vo_2 , other O_2 related parameters such as Feo_2 , Vo_2/hr are also abnormal. After the reconstruction faulty signs disappear as shown in Fig. 9. Therefore we get a conclusion that the 53rd error in the first experiment is transient deviation while from the 67th sampling moment in the second experiment the sensor is malfunction. If the fault is transient we assume it is interfered and use predicted value to replace the wrong data for analysis. If the errors are continuous and several related values are bias we can accurately determine the faulty sensor and isolate it consequently. At the same time the alarm information is sent to base station for further action.

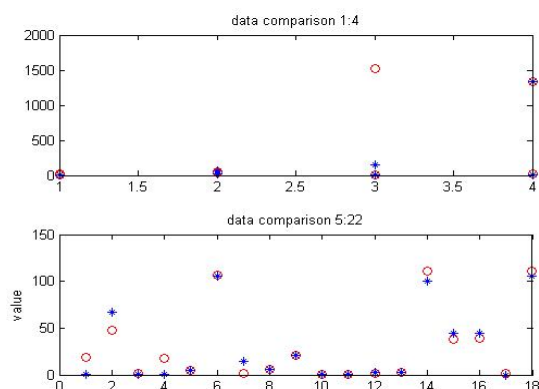


Fig. 8 Comparison measured data and predicted values in continuous errors

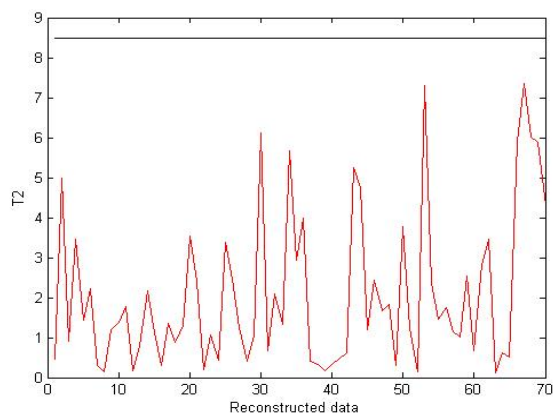


Fig. 9 T^2 Detection with reconstruction of continuous errors

VI. CONCLUSION

Because of the recently research of portable body sensors, using COSMED K4b² to measure consumption of energy in locomotion has got good reputation with a wealth of reports in applications. In order to guarantee the accuracy and availability the reliability is still critical for analysis of physical activity epidemiology and sport performance. We design a comprehensive procedure with statistical techniques of PCA and machine learning method of BP neural network to detect and identify errors from measured data. It is effective and practical proved by simulations. Due to the generalization of this method we will adapt it to the future research in the reliability of body sensor networks.

REFERENCES

- [1] COSMED K4b² Cardio Pulmonary Exercise Testing [Online]. Available: http://www.cosmed.com/images/pdf/productliterature/K4b2_Brochure_EN_C09052-02-93_A4_web.pdf.
- [2] R. Duffield, B. Dawson, H. Pinnington, and P. Wong, "Accuracy and reliability of a Cosmed K4b² portable gas," *Sci Med Sport*, vol. 7, no.1, pp. 11-22, Jan. 2004.
- [3] C. Srl, "K4b² User manual, XVIII Edition," Rome, Italy: <http://www.cosmed.it>.
- [4] W.Haerdle and L. Simar, *Applied Multivariate Statistical Analysis*, Second Edition. Heidelberg: Springer, pp. 37-41, 2007.
- [5] L. H. Xie Tingfeng, Wu Jianjun, "Fault Detection and Diagnosis for Sensors of LRE Based on PCA," *Journal of LRE*, vol. 28, no.1, pp. 668-1672, 1703, Jan, 2007.

- [6] L. Yu, J.-h. Zhu and L.-j. Chen, "Parametric Study on PCA-based Algorithm for Structural Health," in 2010 the Prognostics & System Health Management Conference, pp.1-6.
- [7] P. Picton, *Neural Networks*. New York: Palgrave Houndmills, pp. 215-229, 2000.
- [8] J. Liu, H. Chang, T.Y.Hsu, and X. Ruan, "Prediction of the flow stress of high-speed steel during hot performance using a BP artificial neural network," *Journal of Materials Processing Technology* vol. 103, no.2, pp. 200-205, Feb, 2000.
- [9] Y. Wang, D. Gu, J. Xu, and J. Li, "Back Propagation Neural Network for Short-term Electricity Load Forecasting with Weather Features," in 2009 International Conference on Computational Intelligence and Natural Computing , pp. 58-61.