Open access • Journal Article • DOI:10.4028/WWW.SCIENTIFIC.NET/AMM.110-116.4255

Fault Detection in Tennessee Eastman Process Using Fisher's Discriminant Analysis and Principal Component Analysis Modified by Genetic Algorithm — Source link

Mostafa Noruzi Nashalji, Seyed Mohammad Razeghi, Mahdi Aliyari Shoorehdeli, Mohammad Teshnehlab

Institutions: Islamic Azad University, K.N.Toosi University of Technology

Published on: 01 Oct 2011 - Applied Mechanics and Materials (Trans Tech Publications)

Topics: Kernel Fisher discriminant analysis, Linear discriminant analysis and Principal component analysis

Fault Detection of The Tennessee Eastman Process Using Improved PCA and Neural Classifier

Mostafa Noruzi Nashalji, Mahdi Aliyari Shoorehdeli and Mohammad Teshnehlab

***Abstract*— This paper describes hybrid multivariate method: Principal Component Analysis improved by Genetic Algorithm .This method determines main Principle Components can be used to detect fault during the operation of industrial process by neural classifier. This technique is applied to simulated data collected from the Tennessee Eastman chemical plant simulator which was designed to simulate a wide variety of faults occurring in a chemical plant based on a facility at Eastman chemical.**

***Index Term*— Artificial neural network, Fault detection, Genetic algorithm, Multi layer perceptron, Principal component analysis, Tennessee eastman process.**

I. INTRODUCTION

Analysis of chemical data for detecting faults in chemical process has been intensively studied for the past decade **Error! Bookmark not defined.** ,where a fault is an unpermitted deviation of at least on characteristic property or variable of the system from acceptable/ usual/ standard behavior, and fault detection is a determination of fault present in the system and the time of detection, these definition suggested by the IFAC technical committee safe process.

Data driven technique has been applied in many chemical processes. Chiang et al. applied Fisher Discriminant Analysis (FDA), Discriminant Partial Least Squares (DPLS) and Principal Component Analysis (PCA) for fault detection and diagnosis of Tennessee Eastman Process (TEP) [10]. Russell et al. applied CVA and Dynamic PCA (DPCA) for fault detection in industrial processes [11].

M. N. Nashalji is with the Mechatronics Engineering Group, Electrical Engineering Faculty, K. N. Toosi University of Technology, Tehran, Iran (corresponding author to provide phone: +98 912 6172153; e-email: mostafanoruzi@hotmail.com

M. Aliyari Shoorehdeli is Faculty Member of the Mechatronics Group, Electrical Engineering Faculty, K. N. Toosi University of Technology, Tehran, Iran
m_aliyari@eetd.kntu.ac.ir

M. Teshnehlab is Faculty Member of the Control Group, Electrical Engineering Faculty, K. N. Toosi University of Technology, Tehran, Iran Teshnelab@eetd.kntu.ac.ir

data driven techniques, neural network and PCA for fault monitoring [14]. Zhang used improved kernel Independent Component Analysis (KICA) and SVM to the fault detection and diagnosis in the TEP [*4*]. Modified Discrete Binary Ant Colony Optimization (MDBACO) and Modified Adaptive Chaostic Binary Ant System (ACBAS) are used to reduce the dimensionality of collected data of TEP and SVM as classifier to diagnose the steady faults in [HYPERLINK "file:///E:\\iSamuel\\Work\\Courses\\Fault\\journal-main.docx" \l "Wan06#Wan06" *5*]. In [*7*] , fault diagnosis system is developed by integrating Principal Component Analysis (PCA) with Fuzzy Logic Knowledge-Based (FLKB) systems and this methodology is demonstrated in an academic case study and in the TEP. Xiong et al. used PCA to reduce the dimension of the process data, present a new fault detection method based on artificial immune system for complicated process and illustrated proposed method by the TEP CITATION Xio06 \l 1033 [*6*] . In CITATION Xie06 \l 1033 [*15*] the diagnosis of faults is reduced to a string matching problem and they used an improved Independent Component Analysis based on Particle Swarm Optimization (PSO-ICA) to extract essential variable, the proposed method is illustrated by the application to the TEP challenging process. Akbaryan & Bishno, 2001) used pattern recognition and multi sensor data analysis technique for diagnosing fault of simulated faulty behavior for the TEP CITATION Akb01 \l 1033 [16] . In CITATION Kul05 \l 1033 [17] , a SVM with knowledge incorporation is applied to detect the fault in TEP. Shao & Rong present nonlinear process monitoring based on maximum variance unfolding projections CITATION Sha09 \l 1033 [18] . In CITATION GeZ07 \l 1033 [HYPERLINK "file:///E:\\iSamuel\\Work\\Courses\\Fault\\journal-main.docx" \l "GeZ07#GeZ07" 8] , Ge & Song used ICA-PCA and a new mixed similarity factor for fault detection and diagnosis in the TEP. Xie & Kruger applied PSO-ICA and Support Vector Data Description (SVDD) for fault detection of TEP CITATION Xie061 \l 1033 [19] . In CITATION Zha09 \l 1033 [HYPERLINK "file:///E:\\iSamuel\\Work\\Courses\\Fault\\journal-main.docx" \l "Zha09#Zha09" 9] KICA, Kernel PCA (KPCA) and SVM are used for fault detection and diagnosis in the TEP.

The idea of key dimension selection in classification is select a subset of dimension in which classification result is better or comparable than the result obtained from the full set of dimensions. This paper is used PCA to produce a lower-dimensionality, PCA is a dimensionality reduction technique which transforms correlated original multivariate data to a set of uncorrelated data CITATION Jac91 \l 1033 [20] . The proposed methodology is used GA, SFS and SBS to select Principal Component (PC) of data, which are collected of TEP. Selected PCs are used to make PCA matrix that is called transfer matrix. Neural classifier is used lower dimension reduced data to detect fault. The rest of this paper is organized as follows: TEP is present in section II. In section III PCA introduced. Section IV present GA and the PC selection method based on GA. Section V discusses neural classifier. Section VI shows result and simulation and conclusions coming in section VII

II. TENNESSEE EASTMAN PROCESS

The TEP is a well-known benchmark chemical process, which was firstly introduced by CITATION Dow93 \l 1033 [21] . The TEP has been widely used by the process monitoring community as a source of data comparing various approaches CITATION Rai04 \l 1033 \m KuW95 \m Ger04 \m Him94 \m Chi00 \m Rus00 [HYPERLINK "file:///E:\\iSamuel\\Work\\Courses\\Fault\\journal-main.docx" \l "Chi00#Chi00" 10],[11],[22]-[25] . The TEP was created by Eastman chemical company in open loop operation to provide a realistic industrial process for evaluating process control and monitoring methods CITATION Dow93 \l 1033 [21] .

The process consists of five major units: a reactor, condenser, compressor, separator and stripper. The TEP contain eight components: A, B, C, D, E, F, G and H and produces two products. TEP simulator coded in FORTRAN and a detailed description of the process and simulation is available CITATION Dow93 \l 1033 [*21*] . The simulation code for the process in close loop and the data used for experiment are given by CITATION Mul09 \l 1033 [*26*] . Figure SEQ Figure * ARABIC 1 is a flow sheet for the TEP.

The TEP contains 41 measured and 12 manipulated variables from the controller, all measurements have Gaussian noise. The TEP simulation contains 21 preprogrammed faults as shown in REF _Ref240699312 \h * MERGEFORMAT Table I .

First simulation run was generated with no fault (fault 0) and each of other 20 simulation runs was generated under a different fault (fault 1-21). Sampling time was used to collect the simulated data for the training and testing data is 3 minutes. The simulations started with no faults and the faults were introduced one simulation hour into the run.

III. PRINCIPAL COMPONENT ANALYSIS

PCA is a dimensionality reduction technique CITATION Chi01 \l 1033 [*1*] . PCA is used to linearly project a matrix of data in a low dimensional space, this space spanned by PCs (i.e. eigenvectors corresponding to the large eigenvalues) for the distribution of the training data. PCA determines a set of orthogonal vectors, called loading vector, ordered by the amount of variance explained in the loading vector directions CITATION Chi01 \l 1033 [*1*] .

Let X is the original data set, where each row is a single sample of data set and each column is an observation. Data matrix EMBED Equation.DSMT4 QUOTE consisting of n sample rows and m variable column that are normalized to zero mean and unit variance. The loading vectors are calculated by solving the stationary point of the optimization problem EMBED Equation.DSMT4 QUOTE where EMBED Equation.DSMT4 QUOTE CITATION Chi01 \l 1033 [HYPERLINK "file:///E:\\iSamuel\\Work\\Courses\\Fault\\journal-main.docx" \l "Chi01#Chi01" 1] . PCA and Singular Value Decomposition (SVD) are related, the loading vector can be computed via the SVD, **EMBED Equation.DSMT4** QUOTE (SEQ (* ARABIC 1) where EMBED Equation.DSMT4 QUOTE and EMBED Equation.DSMT4 are unitary matrices, and the matrix EMBED Equation.DSMT4 contains the non-negative (EMBED Equation.DSMT4) and zero offdiagonal elements, the loading vectors are the orthonormal column vectors in the matrix V CITATION Chi01 \l 1033 [HYPERLINK "file:///E:\\iSamuel\\Work\\Courses\\Fault\\journal-main.docx" \l "Chi01#Chi01" *1*] . The variance associated with the *i*th loading vector is equal to the square of the singular value EMBED Equation.DSMT4 .

A reduced set of a smaller number ‘a’ (a<m) of variables, is obtained by a set of loading vectors in the directions (PC) where most of data variation occurs or good for fault detecting. The vector projected into the lower dimensional space, only ‘a’ variables needed to be monitored, as compared with m variables without the use of PCA.

There are several methods for determining ‘a’ (reduction order’s value) CITATION Him94 \l 1033 \m Jac91 \m Qin00 \m Val99 [*20*],[*25*],[*27*],[*28*] . Some techniques for determining ‘a’ and selecting PCs that used in this paper are:

1. The largest singular value,
2. The PRESS statistic and
3. Genetic Algorithm

In the first technique, selecting the columns of the loading matrix EMBED Equation.DSMT4 to correspond to the loading vectors associated with the first ‘a’ singular values.

The dimension of the space can be determined using a cross validation procedure with the Prediction Residual Sum of Squares (PRESS) statistic CITATION Wol78 \l 1033 [*29*] , EMBED Equation.DSMT4 where *i* is the number of subset groups retained to calculated EMBED Equation.DSMT4 and EMBED Equation.DSMT4 is the Frobenius norm. The training set divided into groups. The PRESS statistic for one group is computed based on various dimensions of the space, i, using all the other groups. This is repeated for each group, and the value i associated with the minimum average PRESS statistic determines the dimension of the space CITATION Chi01 \l 1033 [1] . GA introduced in next section.

IV. GENETIC ALGORITHM

To make PCA matrix, the problem is selecting a subset of PC. Finding out what PCs to use in classification task is called PC selection. PC selection is an optimization problem that involves searching the space of PC to find one subset that is optimal to near-optimal with respect to a certain criterion. A number of PC selection (some authors use the term "feature selection" rather than PC selection) method have been proposed in CITATION Das97 \l 1033 \m Blu97 \m Sun03 \m Jai00 \m Yan98 \m Sun04 [29]-[34] . GA provides a simple, general and powerful framework to select good subsets of PCs. GA is problem solving method that has been developed with evolution strategies.

Recently, GAs have attracted more and more attention in the feature -PC- selection area CITATION Sun04 \l 1033 [34] . CITATION Sie89 \l 1033 \m Sun04 [35] and [34] presented one of the earliest studies of GA-based feature selection. In CITATION Yan98 \l 1033 \m YCH98 \m Sun04 \m Cho08 [33], [34],[36] and [37] , GA used for a feature selection. They used binary encoding and standard GA operators to solve the feature selection problem. GA is very efficient technique for PC selection in this paper, because the size of the selection area is enormous (EMBED Equation.DSMT4)

The GA starts with randomly chosen parent chromosomes from the search space to create a population CITATION Sho06 \l 1033 [38] . It works with chromosomes genotype. The population evolves towards the better chromosomes by applying genetic operators modeling the genetic processes occurring in the nature -selection, recombination and mutation CITATION Sho06 \l 1033 [38] .

Sun et al. have employed a simple encoding scheme where the chromosome is a bit string whose length is determined by the number of eigenvectors CITATION Sun04 \l 1033 [34] . Each eigenvector is associated with one bit in the string, if the i^{th} bit is '1', then the i^{th} eigenvector is selected, otherwise, that component is ignored. Each chromosome represents a different subset of eigenvectors.

- Initial population:

The initial population is generated randomly.

- Recombination:

The bit string of two chromosomes, named parent, are cut into two pieces and the information of one side of the bit strings (parents) are swapped, cutting point chosen randomly. This type of recombination is called crossover CITATION Nel01 \l 1033 [39] . Both offspring contains one part of information from one parent and the rest of the other parent.

- MUTATION

Mutation is the other evolutionary operator for GA; its general aim is to create a new individual from only one chromosome by changing one or more genes in it CITATION Sho06 \l 1033 [38] . In this paper, mutation operator just flips a specific bit with a very low probability, it simply inverts on bit in the individual.

- SELECTION

Recombination and mutation operators are made new chromosomes. Selection for replacement creates the new generation from the current one and the obtained offspring. The best N (initial population size) chromosomes in respect of fitness function selected the combined parent-offspring population.

- FITNESS AVALUATION

The fitness evaluation contains two terms: (1) accuracy and (2) the number of feature selected, PCs, CITATION Sun04 \l 1033 [34] .

The PC subset with fewer PCs and better performance is preferred. We used the fitness function shown below to combine the two terms: EMBED Equation.DSMT4 where *miss* represents the misclassification rate of the data samples realized by the neural classifier, N_f is the number of PC chosen by the chromosome, N_{all} is the dimension of PC space and p is a parameter which balanced the misclassification rate and the number of retained feature.

NUERAL CLASSIFIER

A neural network consists of a large number of simple processors called neurons as shown in Figure 2.

Figure 2 illustrates a single layer perceptron with s-neurons. The input vector –in this paper called neural input- is contain r-inputs EMBED Equation.DSMT4 . These inputs are from external source or can come from other units.

The detailed structure of the neurons is shown in Figure 3, neural inputs multiplied by the weight matrix W to form WX. Each neuron has a constant input of 1 shown by x_0, x_0 is multiplied by a bias (w_0) to form w_0x_0. In each neuron WX and w_0x_0 send to the summer. The summer output, n, is $WX+w_0x_0$ (EMBED Equation.DSMT4 QUOTE neurons of single layer perceptron). The output of neuron i is EMBED Equation.DSMT4 , where EMBED Equation.DSMT4 is an activation function of the neurons CITATION Gup03 \l 1033 \m Hag96 [41], [42] . The output of the single layer perceptron is given as EMBED Equation.DSMT4.

Network with several layers as shown in figure 4 has simply cascade three perceptron networks, each layer has its own weight matrix W, its bias vector w_0 and as output vector a. different layers can have different numbers of neurons. A layer whose output is the network output is called an output layer, the other layers are called hidden layers CITATION Hag96 \l 1033 [42] . The networks shown (Figure 4) has an output layer (layer 3) and two hidden layers (layer 1 and 2). Each neuron in the first hidden layer has an input from every neural input with one additional input of 1 -bias-, each neuron in the second hidden layer has an input from every neuron in the first layer additional input like firs layer. In the output layer, there is single neuron called output neuron. The output neuron has an input from every neuron in the second hidden layer and one additional input like first and second layer.

In fault detection application, neural networks used measured and manipulated variables as inputs, while the output represents categories –this neural network called neural classifier- in this paper contains faulty and normal operation. Usually the output is dummy variable ('1' or '0') where '1' indicates an in-class member while '0' indicates a non-class member. Here, dummy variables as output are '1' or '-1' which the margin between the in -class member –shown normal operation- and non-class member-shown faulty operation- is bigger than previous dummy variable.

RESULT AND DISCUSSIONS

Methods used in this paper are:

- Fault detection using all variable by neural classifier.
- Fault detection using PCA based largest singular value with first 15 columns of the loading vector in the matrix V at equation REF _Ref240856596 \h * MERGEFORMAT (1) by neural classifier.
- Fault detection using PCA based PRESS statistic by neural classifier: 400 subset groups of PCs are generated randomly to find good PCs subset by PRESS statistic for fault detecting by neural classifier.
- Fault detection using improved PCA, PCA that used GA to select PC, by neural classifier: GA has several parameters for which no guidance is available on how to specify their values. We used a population size of 100 and 100 generations, with a probability of mutation pm = 0.02 and probability of recombination pc = 0.02 to find near-optimal PC subset.

Neural classifier has three layers, first hidden layer has 25 neurons, second hidden layer has 10 neurons and output layer has one neuron. The activation function of each neuron of hidden layers is hyperbolic tangent sigmoid and output's activation function is linear. Neural classifier is trained using Levenberg-Marquardt learning algorithm CITATION Hag99 \l 1033 [43] . The overall missed detection rate for each method when applied to all disturbances of the testing set and selected PCs are listed in Table II. The sensitivity of fault detection techniques was quantified by calculating the missed detection rate for fault 1-21 of the testing set, for each method. The missed detection rate for all 21 faults as faulty state were computed and tabulated in Table IV. The detection delay for all 21 faults is listed in Table III. The detection delay is recorded as the first time instant in which method exceeded. The numbers in table 2-4 are average of 30 runs of each method.

a. Classification results using all variables

For the TEP data, neural classifier achieved bad detection result by using all variable. This method produced higher missed detection rate than PCA based methods.

b. Classification results with Principal Component selection

The PCA based largest singular value and PRESS statistic performed better than the fault detection using all variable but these methods have high missed detection rate.

It is difficult to compare the GA method with the other methods. Unlike the other methods, this method does not try to find the best subset of a specified size its search space surrounds all the subsets.

The improved PCA with MLP classifier obtained a good accuracy using a few of PCs.

CONCLUSIONS

In fault detection methods, the number of variable or PC greatly affect the ability of fault detection. This paper have focused on performing PC selection on data sets since PC selection is typically done in an off-line mode, the execution time of a particular algorithm is of much less importance than its ultimate classification performance.

This paper illustrates the deservingness of various methods of PC selection to detect fault of TEP by neural classifier.

It was shown that the optimum or near optimum number of PCs, which maximizes the sensitivity of fault detection by neural classifier, can be select by GA. Results show that the proposed method performs good detection capability.

While the improved PCA with neural classifier had the lowest missed detection rate and the lowest detection delay of most faults for the TEP, it should always be used as part of a fault detection strategy.

REFERENCES

[1] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. New York: Springer-Verlag, 2001.

[2] T. Kourti, "Process analysis and abnormal situation detection: from theory to practice," *IEEE Control Systems Magazine*, vol. 22, no. 5, pp. 10-25, Oct. 2002.

[3] J. F. MacGregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process montoring and diagnosis by multiblock PLS method," *American Institute of Chemical Engineering Journal*, vol. 40, no. 5, p. 826–838, 1994.

[4] Y. Zhang, "Fault Detection and Diagnosis of Nonlinear Processes Using Improved Kernel Independent Component Analysis (KICA) and Support Vector Machine (SVM)," *Industrial & Engineering Chemistry Research*, vol. 47, no. 18, pp. 6961-6971, 2008.

[5] L. Wang and J. Yu, "A Modified Discrete Binary Ant Colony

Optimization and Its Application in Chemical Process Fault Diagnosis," in *Simulated Evolution and Learning*, T. .-D. Wang, et al., Eds. New York: Springer-Verlag, 2006, vol. 4247, pp. 889-896.

[6] C. Xiong, Y. Zhao, and W. Liu, "Fault Detection Method Based on Artificial Immune System for Complicated Process," in *Computational Intelligence*, D.-S. Huang, K. Li, and G. W. Irwin, Eds. New York: Springer-Verlag, 2006, vol. 4114, pp. 625-630.

[7] E. Musulin, I. Yélamos, and L. Puigjaner, "Integration of principal component analysis and fuzzy logic systems for comprehensive process fault detection and diagnosis," *Industrial and Engineering Chemistry Research*, vol. 45, no. 5, p. 1739–1750, 2006.

[8] Z. Ge and Z. Song, "Process Monitoring Based on Independent Component Analysis−Principal Component Analysis (ICA−PCA) and Similarity Factors," *Industrial and Engineering Chemistry Research*, vol. 46, no. 7, p. 2054–2063, Mar. 2007.

[9] Y. Zhang, "Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM," *Chemical Engineering Science*, vol. 64, no. 5, pp. 801-811, Mar. 2009.

[10] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 20, no. 2, pp. 243-252, Mar. 2000.

[11] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 81-93, May 2000.

[12] M. Misra, H. H. Yue, S. J. Qin, and C. Ling, "Multivariate process monitoring and fault diagnosis by multi-scale PCA," *Computers and Chemical Engineering*, vol. 26, no. 9, p. 1281–1293, Sep. 2002.

[13] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," *Computers & Chemical Engineering*, vol. 28, no. 8, pp. 1389-1401, Jul. 2004.

[14] J. Chen and C.-M. Liao, "Dynamic process fault monitoring based on neural network and PCA," *Journal of Process Control*, vol. 12, no. 2, pp. 277-289, Feb. 2002.

[15] L. Xie and J. Zhang, "Swarm Intelligent Analysis of Independent Component and Its Application in Fault Detection and Diagnosis," in *Rough Sets and Knowledge Technology*, G. Wang, et al., Eds. Berlin / Heidelberg: Springer, 2006, vol. 4062, pp. 742-749.

[16] F. Akbaryan and P. R. Bishno, "Fault diagnosis of multivariate systems using pattern recognition and multisensor data analysis technique," *Computers & Chemical Engineering*, vol. 25, no. 9-10, pp. 1313-1339, Sep. 2001.

[17] A. Kulkarni, V. K. Jayaraman, and B. D. Kulkarni, "Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process," *Computers & Chemical Engineering*, vol. 29, no. 10, pp. 2128-2133, Sep. 2005.

[18] J.-D. Shao and G. Rong, "Nonlinear process monitoring based on maximum variance unfolding projections," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11332-11340, Oct. 2009.

[19] L. Xie and U. Kruger, "Statistical Processes Monitoring Based on Improved ICA and SVDD," in *Intelligent Computing*, D.-S. Huang, K. Li, and G. W. Irwin, Eds. Springer Berlin / Heidelberg, 2006, vol. 4113, pp. 1247-1256.

[20] J. E. Jackson, *A users guide to principal components analysis*. New York: Wiley, 1991.

[21] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245-255, Mar. 1993.

[22] A. Raich and A. Çinar, "Statistical process monitoring and disturbance diagnosis in multivariable continuous processes," *American Institute of Chemical Engineers (AIChE)*, vol. 42, no. 4, pp. 995-1009, 2004.

[23] W. Ku, R. H. Storer, and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 30, no. 1, pp. 179-196, Nov. 1995.

[24] J. Gertler, W. Li, Y. Huang, and T. McAvoy, "Isolation enhanced principal component analysis," *American Institute of Chemical Engineers (AIChE)*, vol. 45, no. 2, pp. 323-334, Apr. 2004.

[25] D. M. Himes, R. H. Storer, and C. georagkis, "Determination of the Number of Principal Components for Disturbance Detection and Isolation," *In Proceedings of the american control conference*, pp. 1279-1283, Jun. 1994.

[26] Multiscale Systems Research Laboratory. [Online]. HYPERLINK "http://brahms.scs.uiuc.edu/" http://brahms.scs.uiuc.edu

[27] S. J. Qin and R. Dunia, "Determining the number of principal components for best resconstruction," *Journal of Process Control*, vol. 10, no. 2, pp. 245-250, 2000.

[28] S. Valle, W. Li, and S. J. Qin, "Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods," *Industrial and Engineering Chemistry Research*, vol. 38, no. 11, pp. 4389-4401, 1999.

[29] S. Wold, "Cross-validatory estimation of the number of components in factor and principal components models," *Technometris*, vol. 20, pp. 397-405, 1978.

[30] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131-156, 1997.

[31] A. L. Blum and P. Langley, "Selection of Relevant Features and Examples in Machine Learning," *Artificial Intelligence*, vol. 97, pp. 245-271, 1997.

[32] Z. Sun, G. Bebis, and R. Miller, "Boosting Object Detection Using Feature Selection," in *IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03)*, Jul. 2003, pp. 290-296.

[33] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37, Jan. 2000.

[34] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 2, pp. 44-49, 1998.

[35] Z. Sun, G. Bebis, and R. Miller, "Object detection using feature subset selection," *Pattern Recognition*, vol. 37, no. 11, pp. 2165-2176, Nov. 2004.

[36] W. Siedlecki and J. Sklansky, "A note on genetic algorithm for large-scale feature selection," *Pattern Recognition Letters*, vol. 10, p. 335–347, 1989.

[37] C. Y., B. D., and B. D., "Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision," *Journal of the science of food and agriculture*, vol. 76, pp. 77-86, 1998.

[38] H.-W. Choa, et al., "Genetic algorithm-based feature selection in

high-resolution NMR spectra," *Expert Systems with Applications: An International Journal*, vol. 35, no. 3, pp. 967-975, Oct. 2008.

[39] E. G. Shopova and N. G. Vaklieva-Bancheva, "BASIC—A genetic algorithm for engineering problems solution," *Computers and Chemical Engineering*, vol. 30, no. 8, pp. 1293-1309, Jun. 2006.

[40] O. Nelles, *Nonlinear system identification*. Berlin: Springer, 2001.

[41] M. M. Gupta, L. Jin, and N. Homma, *Static and dynamic neural networks: from fundamentals to advanced theory*. Wiley-IEEE, 2003.

[42] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*. Boston: PWS Publishing, 1996.

[43] M. T. Hagan and M. Menhaj, "Training feed-forward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989-993, 1999.