# Data-driven fault diagnosis-tolerant control integrated technique for solid oxide fuel cell

Yilin Wang<sup>a,b</sup>, Shuanghong Li<sup>a,b\*</sup>, Yupu Yang<sup>a,b</sup>

<sup>a</sup> Department of Automation, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, PR China <sup>b</sup> Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai 200240, PR China

## Abstract

Fault monitoring and diagnosis system are very important in detecting system failures and keeping the stability of production line for the modern industrial system. This paper studies the data-driven fault diagnosis and tolerant control integrated technique focused on Solid Oxide Fuel Cell (SOFC) system, especially for the stack degradation fault. The multivariable statistical approach Support Vector Machine (SVM) as well as Principal Component Analysis (PCA) are studied for the multi-fault classification and diagnosis purpose. Then based on the diagnosis results, the decision-making part is designed to select appropriate fault reconfigurable control strategy which can handle five types of stack faults and recover the thermal and electrical parameters to normal operating condition. The core of integrated data-based fault monitoring and control approach is to take full advantage of available SOFC measurements data aiming to acquire the useful fault condition information and choosing adequate fault recovery control method. The results are that the proposed strategy can keep the system power near the normal operating state and make the SOFC temperature steady near the enactment value when working in faulty conditions, which may result in both lifetime and durability improvement for the SOFC system.

Keywords: Fault diagnosis, SOFC, fault tolerant control, PCA, SVM

# 1. Introduction

Hydrogen-fueled solid oxide fuel cells (SOFC) system can directly generate electric power from hydrogen; the generated electric power has high operation temperature, zero greenhouse gas emissions, and hydrogen impurity tolerance [1-4]. The SOFC system is especially suitable to structure distributed alternative power station since it can produce electricity with very high efficiency. Moreover, the quiet operation characteristic associated with reduced emissions feature make it appropriate to locate SOFC systems in the downtown areas and densely populated regions for commercial and residential loads. Although SOFC shows its high-energy conversion efficiency and environmental compatibility, the electricity plant based on stack of fuel cells still suffer from short lifetimes and low reliabilities. Therefore, the development of systems for monitoring SOFC system, the especially ones using automatic online diagnosis and fault tolerant control, is critical for the progress of the SOFC area.

Timely and precise fault detection and fault diagnosis for the underlying procedure can lessen harmfulness, improve the safeness and reliability of the equipment manipulations, and yet decrease manufacturing costs. The SOFC system (Fig. 1) fault diagnosis includes the following two kinds of methods (1) the physical principles [5, 6], and (2) data-driven approach based on history data [7]. The main tasks of fault diagnosis are fault detection, fault classification, fault location, fault recovery and so on. As soon as the fault is detected, the fault class then needs to be identified.

<sup>\*</sup> Manuscript received August 21, 2019; revised December 10, 2019.

Corresponding author. Tel.: +8613761268192; *E-mail address*: lishuanghong@sjtu.edu.cn. doi: 10.12720/sgce.9.2.357-365

However, since the industry processes become more and more complicated, it is almost impossible to build a precise model relying only on the physical principles of the systems. With the development of the automation degree, a large number of sensors and actuators installed in the modern plants generate large numbers of process data. Modern industries have embraced the dawn of a data-based approach due to the extreme difficulty in obtaining the physical models for the complicated processes. Compared with the well-developed model-based approaches, data-based techniques provide different alternative solutions for different industrial issues under various operating conditions, such as system monitoring and control. The large size of data makes great challenges not only to capture, manage, and store large data sets but also to efficiently interpret the information embedded in the process measurement via data mining techniques. Data mining technique has become a potential solution to find out useful information to help researchers make appropriate decisions [7].



Fig. 1. SOFC structure with anode offgas recycle

The purpose of the fault diagnosis is not just to know the mode of the fault in plants, but the more important task is to avoid, repair, and even recover the plants due to the diagnosed results. Although many researchers study the fuel cells fault diagnosis while others study the fault tolerant control. However, combing the fault diagnosis with the tolerant control is still very lacking. Wu [8] designs a fault tolerance control of SOFC systems based on nonlinear model predictive control. The proposed fault tolerant control strategy can track the voltage and make the SOFC temperature steady near the enactment value while working in the faulty conditions. However, the model-based fault diagnosis part largely depends on the model accuracy and the model building needs too much time. Sun [9] presents a hierarchical structure for fault detection and fault-tolerant control of SOFC stack, but the fault detection is based on linear state observer which is impossible to observe complex nonlinear relationship in the whole SOFC system.

To overcome the limitations presented above, data-driven fault diagnosis-tolerant control integrated technique for a complete SOFC system is proposed, which consists of a fault diagnosis part, a decision-making part and five backup controllers to solve the following problems:

(1) Data-driven Fault diagnosis.

(2) Fault tolerant control.

The advantages of the proposed strategy are as follows: 1) The diagnosis method combines fault detection with classification, because the normal condition is treated as one type mode, which does not need to select fault threshold; 2) The strategy combines diagnosis with fault tolerant control, and 3) the training process of the classification is offline, but the diagnosis can be online, which is very rapidly for the online control usage.

#### 2. Data-driven Diagnosis-recovery Integrated Techniques

#### 2.1. Data-driven diagnosis framework

Generally, a data-driven fault diagnosis system consists of three steps [8]. The first step is data collection with sensors and controllers. Then data processing includes normalization, dimension reduction, and statistical feature extraction is executed in the next step. Some feature selection techniques including PCA is commonly used to reduce the dimension of the recorded data set to save computation load [10]. In the final step, the operation status of the monitored process is predicted by maintenance personnel manually or by some automated computational intelligence methods. Several classification techniques have been proposed and greatly improved over the past few years, among them, SVM is an important

classification approach and the most common used method. It can handle the classification problem with finite samples and large feature space. SVM is preferred here in SOFC system data identification because it is a presentative nonlinear classifier and its superior generalization ability [7, 10].

## 2.2. Principal Component Analysis (PCA)

The PCA, which has been used to extract the principal components from the data set, is introduced in this section. It is capable of making interrelated variables uncorrelated via singular value decomposition (SVD) or eigenvalue decomposition approach. PCA can extract main factors from multiple variable vectors which can reveal the characters of them. The central advantage of the PCA is to decrease the dimensionality of the data set, while reserving the main variation factor as much as possible. In that way, the calculation burden is largely decreased by transforming the high-dimensional data into low-dimensional data.

In the production line, directly obtaining data features from sensors traditionally lead to an overlapping pattern. In order to reduce the redundant information, PCA is performed for feature extraction. It has been employed to reduce the data dimensionality and to transform correlated variables into independent ones. In our work, the PCA function of dimensionality reduction is applied, the algorithm is as follows.

Table 1. PCA algorithm

Input: data set  $D = \{x_1, x_2, \text{to } x_m\}^T$ Low-dimensional space dimension: NNormalizing: For each variable  $\overline{x_i} = x_i - \frac{1}{n} \sum_{i=1}^n x_i$ End for  $X = \{\overline{x_1}, \overline{x_2}, \text{to } \overline{x_m}\}^T$ Covariance matrix:  $C_x = XX^T$ Eigenvalue decomposition:  $C_x = W^T \Lambda W, \ \Lambda = diag[\lambda_1, \text{to } \lambda_n]$ Principle analysis: Selecting top N largest  $\lambda$ and the corresponding eigenvector  $w_1, w_2$ , to  $w_N$ Output: projection matrix:  $W_p = (w_1, w_2, \text{to } w_N)$ 

## 2.3. Support Vector Machine (SVM)

Statistical learning theory is a discipline which studies the law of machine learning under limited training sample. It provides a new classification method named support vector machine [7, 10]. This method is established on the basis of the VC dimension theory and structural risk minimization principle of the statistical learning theory. Support vector machine has many advantages in handling the small sample, non-resistance, and high dimensional problems, making it an excellent machine learning algorithm. It is applied to process classification problem, regression problem, and other learning tasks as a popular machine learning method. Further details of SVM are discussed below. *2.3.1. SVM* 

Taking into account a training sample classification problem, suppose that the training data set  $(x_1; y_1), \dots, (x_q; y_q), x_i$  is the *i*th sample of input samples.  $y_i$  is the output sample corresponding to the *i*th input sample. SVM [7, 10] needs to solve the coming primal issue:

$$\min_{w,d,\zeta} \frac{1}{2} w^{\mathrm{T}} w + C \sum_{i=1}^{q} \zeta_{i}$$
s.t.  $y_{i} (w^{\mathrm{T}} \Phi(x_{i}) + d) \ge 1 - \zeta_{i}$ 

$$\zeta \ge 0 \qquad i = 1 \dots q$$
(1)

where w is a q-dimensional vector, C is the penalty parameter,  $\zeta_i$  is a measure of distance between the misclassified point and the separating hyperplane,  $\Phi(x_i)$  projects the original data sets into a higher dimensional space and d is a scalar.

## 2.3.2. Kernel

Support vector machine is originally used to process linearly separable cases. But not all the problems can be linearly separated. When meeting with non-linear cases, this linear classifier may not work well. Therefore, a method called kernel was introduced to handle non-linear issues using linear classifier.

Based on the pattern recognition theory, low-dimensional space non-linear inseparable model turns into linear separable by nonlinearly mapping it into a high-dimensional feature space. However, if we use this technology in high-dimensional space classification or regression directly, it can be difficult to determine the form, the parameters and the feature dimensional of the nonlinear mapping function. Using kernel technology can effectively solve this problem. There are several commonly used kernel functions listed as follows [10]:

1) Linear

$$k(x_i, x_j) = x_i^{\mathrm{T}} x_j \tag{2}$$

2) Polynomial

$$k(x_i, x_j) = (x_i^{\mathrm{T}} x_j)^{\mathrm{d}}$$
(3)

3) Radial basis function (RBF)

$$k(x_i, x_j) = \exp(-||x_i - x_j||^2) / \delta^2$$
(4)

where c, k and v are constants, b is the degree of the polynomial,  $\delta$  is the width of RBF kernel.

Due to the nonlinear characters of the SOFC data, the RBF kernel function is selected. Mostly RBF is a reasonable choice because of the following reasons:

1) RBF non-linearly maps the samples to a high dimensional space, i.e., it can handle the cases where the relation between class labels and its attributes is non-linear.

2) The number of hyperparameters which affect the complexity of model is less for RBF kernel.

## 2.4. Data-driven fault classification

In the data-driven fault classification process, two kinds of data sets should be prepared first: the training data set and the testing data set. The training data set should be labeled, which means the fault type of the data should be known then the data can be used to train the classifier. The high-dimensional data should be reduced to low-dimensional which is much faster for the data processing. Due to the data is redundant, the feature extracted by PCA will not lose its primary feature. The low-dimensional is used to train the multi-class SVM, and then the trained SVM is used to classify the fault type of the testing data.

The implementation of the selected control variable pairings and the decoupling design are addressed

in the following sections, shown in Fig. 2. The relationship of the control inputs and the outputs, the integrated decoupling control scheme is due to our previous work [11].



Fig. 2. Integrate fault diagnosis-tolerant control

## 3. Simulation Results

In Simulink environment of MATLAB 2017a software, the proposed data-driven fault diagnosistolerance control integrate technique strategy for the SOFC system is given. The data-driven PCA and SVM are running in the Python 2.7 and then the .py document can also running in the MATLAB software and transmit classification results to the SOFC controller. The KW-scale SOFC stand-alone system is developed as the platform for the data-driven fault diagnosis-tolerant control integrated technique for Solid Oxide Fuel Cell. Special attention is drawn on the system-level thermal and electrical parameter simulation. The SOFC system is modeled based on transportation and conservation principles. The model has been built through many previous efforts [5, 11]. Given that the focus of this study is system controller design, the dynamic modeling method is briefly introduced. The model runs in a MATLAB/Simulink platform on a computer with 4.4GHz and 16 G memory. The kW-scale SOFC standalone system model comprises a planar SOFC stack, a burner, and two heat exchangers (Fig. 2), in which a special consideration for stack spatial temperature management is conducted by an air bypass manifold around heat exchangers [3, 4]. Particularly, an AOR structure is designed in this system as a fault tolerant structure [11].

#### 3.1. Data preparing for the SOFC

The operation condition for the SOFC system includes two states, the normal operation state, and the fault state. SOFC stack degradation [7]: The performance of a fuel cell stack is strongly influenced by its internal sources of voltage losses (activation, Ohmic resistance and diffusion losses), which can be described by the relationship between the voltage and the current density [7, 8]. In this study, stack degradation is calculated by increasing the overall stack resistance to its nominal value.

$$r_f = (1+\lambda) \cdot r \tag{5}$$

where  $\lambda$  is the stack fault sensitive factor. If the stack is normal,  $\lambda$  is 0. Whereas if the stack occurs fault,  $\lambda$  is greater than 0.

The normal operation condition is about: the system power is about 5000W, and the stack temperature is about 960K, the burner temperature is about 1140 K.

The errors are set by power decrease due to the stack degradation is about 2%, 4%, 6%, 8% and 10% respectively. Due to the control target is to recover the degradation by increasing the fuel and keep the burner temperature steadily, the stack degradation degree should not be set too much.

After setting the error magnitude, a data set is created by 201 variables, and each variable collected 4000 points. Then the training data of each operating condition is about 200 ×4000 matrix.

#### 3.2. PCA-SVM fault diagnosis



Fig. 3. Integrate fault diagnosis-tolerant control

It is very important to reduce the data dimension before applying SVM. Avoiding attributes with greater numerical range dominating those with smaller numerical range is the primary advantage of scale. Avoiding numerical difficulties during the calculation is the other advantage. Therefore, attributes with large values might lead to numerical problems. The PCA is applied to decrease the dimensionality of the data from 201×4000 to 8×4000. Principal Component Variance percent explained is shown in Fig. 3. In this study, we select 8 largest principal components and the retained variance ratio is 99.10%.



Fig. 4. Fault classification results

The Fig.4 shows the testing results of the training data. For the first 2434 points, the SOFC is running in the normal operation state and then fault 1,2,3,4 data is collected each about 1784 points. One can see from the Fig.4, the multi-class SVM can classify the fault type, but an obvious phenomenon can be seen from the classification results that the fault will be detected as previous faults at the very beginning. For

example, the fault mode 5 is classified as the mode 1 to mode 4 at the start period. That because the stack fault about 10% do not occur abruptly. In fact, the 10% stack degradation should occur slowly from 2%, 4%, 6%, 8% and finally reach 10%.

Stack Fault	Fault class	Classification accuracy without PCA	Classification accuracy with PCA
Normal	Mode 0	99.9%	99.9%
2%	Mode 1	96%	96%
4%	Mode 2	90%	92%
6%	Mode 3	85%	91%
8%	Mode 4	79%	91%
10%	Mode 5	75%	88%
PCA+ Training+ Testing		126.5s	10.4s

Table 2. Fault diagnosis accuracy

The fault diagnosis accuracy is shown in TABLE IV. From the results, it indicates that the classification with PCA performs better than the one without PCA and the running time of the data-driven fault diagnosis with PCA is 91.78% less than that without it. The reason lies on three aspects.

1) Firstly, after reducing the dimension, the number of features is reduced, so overfitting can be avoided effectively.

2) Secondly, PCA denoises the disturbance variables and obtains the most valuable features.

3) Thirdly, the extracted features can largely improve the training and testing speed of the classifier.

Therefore, the accuracy and the rapidity of the multi-class SVM are improved by using the PCA. The

fault classification results will be used in the next section to depend on which fault recovery controllers to select.

#### 3.3. Fault accommodating control

This section performs fault tolerant control depends on the fault diagnosis results obtained in the last section.

All of the SISO controllers in the decoupling control scheme are PID-based controllers due to their simple structure and good performance in engineering practice with robustness and stability. The control parameters of the decoupling control system are tuned by the Ziegler–Nichols tuning method which is suitable for engineering practice [11].

To test the validation of the proposed control strategy in the case of the normal, the power load is varied from 0 to 4950W and kept by the nominal control for 40000s. The burner temperature is kept at about 1141K and the stack temperature is kept at 1075K. The control actuations such as air flow rate, fuel flow rate and the fuel rate to burner are maintained at constant value shown in Fig.5.



Fig.5. Stack fault and diagnosis results: (a) Ohmic resistance multiple (b) PCA-SVM fault classification results

To investigate the performance of the proposed data-driven fault diagnosis and tolerant control technique, the stack fault about 10% Ohmic resistance increase is simulated. At the first 20000s the SOFC is operated at normal condition, then the stack fault occurs at t=20000s. The Ohmic resistance multiple is varied from 1 to 1.1 in about 200s. The data-driven fault diagnosis part detect the fault and classify the fault type from 0 to 1, 2, 3, 4, 5 respectively with deepening of the stack fault shown in Fig.5.(b). Furthermore, the reconfiguration algorithm decides to activate corresponding recovery strategy.



Fig.6. SOFC response with and without fault diagnosis and fault tolerant control under Fault 5: (a) Variations of system power (b) Variations of burner temperature (c) Variations of stack temperature (d) Variations of air flow rate (e) Variations of fuel flow rate (f) Variations of fuel flow rate to burner

Fig. 6.a-c show the comparison of SOFC power, burner temperature and stack temperature with and without fault tolerant control. The discrepancies between the two control modes indicate that the fault occurs during the nominal control period can cause power coastdown and the unstability of the temperature which may result in both lifetime and durability decrease. The fault tolerant control can attenuate the effect of the fault and keep the burner and stack temperature at the desired value by the control actuations shown in Fig.6 d-f. The power can return to about 4950W by the control actuations which is about at the initial value.

The Fig. 6 d-f present the air flow rate, fuel flow rate and the fuel flow rate to the burner respectively. The sudden changes in the flow rates are required by the tolerant controller to make up for the stack fault.

# 4. Discussion and Conclusion

This study proposes data-driven fault diagnosis and tolerant control for SOFC system. The machine learning method PCA and SVM are imported to classify the fault type based on the data collected from

the SOFC system. The fault tolerant control is designed by six different backup controllers to handle different types of operating conditions which can be selected by the PCA-SVM fault classifier. The results of the simulation indicate that the proposed strategy can keep the system power at the desired value while maintaining the SOFC burner and stack temperature at both normal and faulty conditions. In this way, both lifetime and durability of material in SOFC can be largely improved and the power supply stabilization can be guaranteed.

#### **Conflict of Interest**

The authors declare no conflict of interest.

#### Acknowledgements

This work is partly supported by National Nature Science Foundation of China (No. 61273161 and No. 51777122).

#### References

- Lee JH, Jo JS, Choi SW, and Han SB. A 10-kW SOFC low-voltage battery hybrid power conditioning system for residential use. *IEEE Trans. Energy Convers.*, 2006; 21(2): 575-585.
- [2] Colson CM and Nehrir MH. Evaluating the benefits of a hybrid solid oxide fuel cell combined heat and power plant for energy sustainability and emissions avoidance. *IEEE Trans. Energy Convers.*, 2011; 26(1): 140-148.
- [3] Cao HL, Li X, Deng ZH, Li J, and Qin Y. Thermal management oriented steady state analysis and optimization of a kW scale solid oxide fuel cell stand-alone system for maximum system efficiency. *Int J Hydrogen energy*, 2013; 38: 12404-12417.
- [4] Mueller F, Jabbari F, and Gaynor R. Novel solid oxide fuel cell system controller for rapid load following. J. Power Sources, 2007; 172: 308-323.
- [5] Greco A, Sorce A, Littwin R, Costamagnac P, and Magistri L. Reformer faults in SOFC systems: experimental and modeling analysis, and simulated fault maps. *Int J Hydrogen Energy.*, 2014; 39(36): 21700-21713.
- [6] Comminges C, Fu QX, Zahid M, Yousfi NS, Bucheli O. Monitoring the degradation of a solid oxide fuel cell stack during 10,000h via electrochemical impedance spectroscopy. *Electrochimica Acta.*, 2012; 59: 367-375.
- [7] Costamagna P, Giorgi AD, Magistri L, Moser G, Pellaco L, and Trucco A. A Classification Approach for Model-Based Fault Diagnosis in Power Generation Systems Based on Solid Oxide Fuel Cells. *IEEE Trans. Energy Convers.*, 2016;31(2): 676-687.
- [8] Wu XJ, and Gao DH. Fault tolerance control of SOFC systems based on nonlinear model predictive control. Int J Hydrogen energy, 2017; 42(4): 2288-2308.
- [9] Sun YL, Ghantasala S, El-Farra NH. Monitoring and fault tolerant control of distributed power generation: application to solid oxide fuel cells, 2010 American Control Conference, Baltimore, USA, 2010; 448-453.
- [10] Jing C, and Hou J. SVM and PCA based fault classification approaches for complicated industrial process. *Neurocomputing*, 2015; 167: 636–642.
- [11] Li SH, Gong L, and Yang YP. Fault tolerant control of an anode offgas recycle based SOFC system. 2017 American Control Conference, Seattle, USA, 2017; 4117-4122.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.