Feature Selection based on PCA and the Transient Identification in Nuclear Power Plants

Shuqiao Zhou, Chao Guo, Xiaojin Huang
Institute of Nuclear and New Energy Technology
Collaborative Innovation Centre of Advanced Nuclear Energy Technology
Key Laboratory of Advanced Reactor Engineering and Safety of Ministry of Education
Tsinghua University, Beijing 100084, China
zhousq@tsinghua.edu.cn; guochao@tsinghua.edu.cn; huangxj@tsinghua.edu.cn

ABSTRACT

For the sake of enhancing safety and achieving more economic benefits, it is very important to timely and correctly identify the transients during the operation of nuclear power plants. There are thousands of monitoring signals in a nuclear power plant. It is difficult to complete the calculations timely if all the related signals are chosen for an algorithm of transient identification. Moreover, the irrelevant state variables may also do harm to the correctness of the identification process. Thus, one of the key tasks for transient identification is to figure out effective methods to extract the features from state variables. In this paper, we will address an efficient method for feature selection and transient identification based on principal component analysis (PCA). The effectiveness of the method are illustrated by the tests based on the data from the simulator designed for the High Temperature gas-cooled Reactor Pebble-bed Module (HTR-PM), which is under construction in China and the construction will be completed around 2017.

Key Words: feature selection, transient identification, principal component analysis

1 INTRODUCTION

Nuclear power plants are complex systems, as they contain many kinds of devices and relevant processes. Compared with other industrial systems, safety is especially important for nuclear power plants. Several operators are assigned and the monitoring systems are designed to supervise the whole plant [1]. It is the first step– and also a very important step– to identify what kind of abnormal event was happening, as the operators can make the right decision only if they have determined what was happening. Moreover, when an abnormal event has happened and turned to be an accident, the number of parameters of the systems and alarms might be too big to be dealt with timely [2][3]. Thus, it is very important and beneficial to identify the abnormal events before they turn into accidents.

In a nuclear power plant, there are thousands of measured variables from field devices (i.e. sensors, valves and pumps) and variables synthesized by the controlling stations for the process control. If we directly use all relevant variables for transient identification, the number of the variables will be too large and the computation time will not be acceptable. In this paper, we propose a method to use PCA to handle this problem. By the proposed method, the sample data of the related variables can be pre-treated by PCA. Then the dimensionality can be reduced and computation consumption can be significantly cut down.

In the remainder of this paper, we first introduce how to reduce the dimensionality of the samples of the variables related to specific transients. Then how to identify the transients based on samples with reduced dimension is introduced. In section 4, the proposed method is verified by sample data from the plant simulator of HTR-PM [4]. At last, we draw a conclusion in section 5.

2 STATE VARIABLE SELECTION BASED ON PCA

In HTR-PM, there are several thousands of monitoring variables, such as the measured temperature, rate of flow, pressure. In one hand, for a certain failure, the related process would only affect some of the relevant measures, rather than changing all these measures. In the other hand, we cannot use all measures in a transient identification method, as it is very time consuming for the computing with a high dimensionality or the results will be even worse by the influence of the data from irrelevant variables. Typically, PCA, canonical variate analysis (CVA) and partial least squares (PLS) are used for features extraction and dimensionality reduction [5][6]. In this section, how to extract the principal components from the sample data based on PCA is introduced.

Suppose there are m sensors used to monitor the running state of the plant. At time t, a sample $\boldsymbol{v}(t)=\left[v_1(t),v_2(t),\cdots,v_{\hat{m}}(t)\right]^T$ can be obtained by these $\hat{m}$ sensors. Suppose k samples are selected to be considered for each specific transient, the matrix of the samples can be obtained as

$$\hat{T}_i=\begin{bmatrix} v_1(1) & v_1(2) & \cdots & v_1(k) \\ v_2(1) & v_2(2) & \cdots & v_2(k) \\ \vdots & \vdots & \ddots & \vdots \\ v_{\hat{m}}(1) & v_{\hat{m}}(2) & \cdots & v_{\hat{m}}(k) \end{bmatrix} \in R^{\hat{m}\times k}. \tag{1}$$

In $\hat{T}_i$, each row represents the measures of a state variable at different k time points and each column denotes a sample of $\hat{m}$ variables at a time. If there are n transients to be considered, the related matrix can be formularized as

$$X=\left[\hat{T}_1,\hat{T}_2,\cdots,\hat{T}_n\right]. \tag{2}$$

It is necessary to mention that the data obtained from sensors are normalized to a range from 0 to 1 based on the maximal value and minimal value of each variable before using them to form X.

The covariance of variables can be calculated as

$$C=\operatorname{cov}(v)\approx\frac{1}{\hat{m}}XX^T. \tag{3}$$

By using eigenvalue decomposition, the results can be obtained as follows.

$$E^TCE=\Lambda=\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_{\hat{m}} \end{pmatrix}. \tag{4}$$

The rows in E^T are the eigenvectors related to $\lambda_i, i\in\left[1,\hat{m}\right]$. Suppose $\lambda_1>\lambda_2>\cdots>\lambda_{\hat{m}}$, the first m ($m<\hat{m}$) rows of E^T can be selected to form a new matrix P. Then, P can be used as a projection matrix to select the principal components of X. Each transient can be represented as follows after the projection.

$$T_i = P\hat{T}_i = \begin{bmatrix} v_1(1) & v_1(2) & \cdots & v_1(k) \\ v_2(1) & v_2(2) & \cdots & v_2(k) \\ \vdots & \vdots & \ddots & \vdots \\ v_m(1) & v_m(2) & \cdots & v_m(k) \end{bmatrix} \in R^{m\times k}, \mathrm{m} < \hat{\mathrm{m}}. \tag{5}$$

3 TRANSIENT IDENTIFICATION BASED ON PRINCIPAL COMPONENTS

According to [7] , T_i in (5) represents a specific transient. Suppose $\hat{y} = \boldsymbol{v}(t_{current})$ denotes the current state that needs to be identified and $y = P\hat{y}$ will be used in following identification algorithm. By rewriting the elements, the matrix T_i can be described as following vector:

$$A_i = \left[v_1(1), v_1(2), \cdots, v_1(k), v_2(1), v_2(2), \cdots, v_2(k), \cdots, v_m(1), v_m(2), \cdots, v_m(k)\right]^T. \tag{6}$$

Suppose there are n transients to be identified, the set of the related $A_i, i \in [1,n]$ can be represented by a matrix as:

$$\hat{A} = \left[A_1, A_2, \cdots, A_n\right]_{mk\times n}. \tag{7}$$

Suppose there is a transient y to be identified, the identification can be formulized as

$$y = x_1 A_1 + x_2 A_2 + \cdots + x_n A_n = \hat{A}x, \tag{8}$$

where $x = \left[x_1, x_2, \cdots, x_n\right]^T$ is the coefficient vector.

Suppose the number of classes of transients is N_c and the related classes are $C_i, i \in [1, N_c]$. If the unknown $y \in C_i$, the sum of coefficients related to C_i will be much larger than the sum of the remained ones:

$$\sum_{k:A_k \in C_i} x_k \gg \sum_{k:A_k \in C_j, j\neq i} x_k. \tag{9}$$

Thus, we can obtain the class of y by solving (8) first and then calculating the sum of the coefficients related to each class respectively. Finally, we pick the class with the largest coefficient sum as the class of unknown y. The coefficients $x_i, i \in [1,n]$ can be obtained as

$$x = \hat{A}^{\dagger} y, \tag{10}$$

where $\hat{A}^{\dagger}$ denotes the Moore-Penrose inverse of $\hat{A}$. When $\hat{A}$ is undetermined ($mk < n$), we can refer to the fast algorithm, which is called SL0, for sparse decomposition in [8]. Compared with the discussion in [7], in this paper the movable windows are not considered and $\hat{A}$ has $mk > n$, thus the coefficients are obtained by (10) directly.

4 VERIFICATION

4.1 Sample data

As Table I shows, 14 typical transients related to the postulated accidents in HTR-PM are selected to verify the proposed method. These transients are key events related to the safety of the reactor and whole power plant. Together with the normal state T_1, 15 cases are from five classes (C_1~C_5). T_1~ T_{11} are used to build matrix $\hat{A}$ and the remained T_{12}~ T_{15} are used to verify the identification capability of the method proposed in this paper.

There are many state variables related to these transients while the five most relevant state variables are chosen here: reactor heat power, inlet helium temperature, outlet helium temperature, feed water flow and helium flow. The data of these transients are from the plant simulator, since HTR-PM is still under construction.

During the simulation, it is supposed that the data from the sensors are sampled with an interval of 1 s and the total simulation time period is 300 s. In the beginning, the plant works in normal state and the transients happen at 45 s. Because of space limitations, here we just take T_3 and T_8 for examples. The curves of the variables are shown in Figure 1. Accordingly, the related parameters referred above are: $\hat{m}=5$, $k=300$. Thus, without the process of dimensionality reduction by PCA (i.e. all these five variables are directly used to form $\hat{A}$), the dimension of $\hat{A}$ is 1500×11.

Table I. Typical transients of HTR-PM

Transients	Class	Name
T_1	C_1	normal status
T_2	C_2	control rod ejection accident
T_3	C_2	control rod ejection accident (90% FP)*
T_4	C_2	control rod ejection accident (50% FP)*
T_5	C_3	control rod ejection accident without scram
T_6	C_3	control rod ejection accident without scram (90% FP)*
T_7	C_3	control rod ejection accident without scram (50% FP)*
T_8	C_4	water ingress accident (at position 1 (p1))
T_9	C_4	water ingress accident (at position 2 (p2))
T_{10}	C_5	water ingress accident (p1) with relief-pressure failure
T_{11}	C_5	water ingress accident (p2) with relief-pressure failure
T_{12}	C_2	control rod ejection accident (95% FP)*
T_{13}	C_3	control rod ejection accident without scram (95% FP)*
T_{14}	C_4	water ingress accident (at position 3 (p3))
T_{15}	C_5	water ingress accident (p3) with relief-pressure failure

*FP represents full power level (e.g., 90% FP indicates at 90% of full power level). If there is not a mark of start, it means that the case is running at 100% power level.

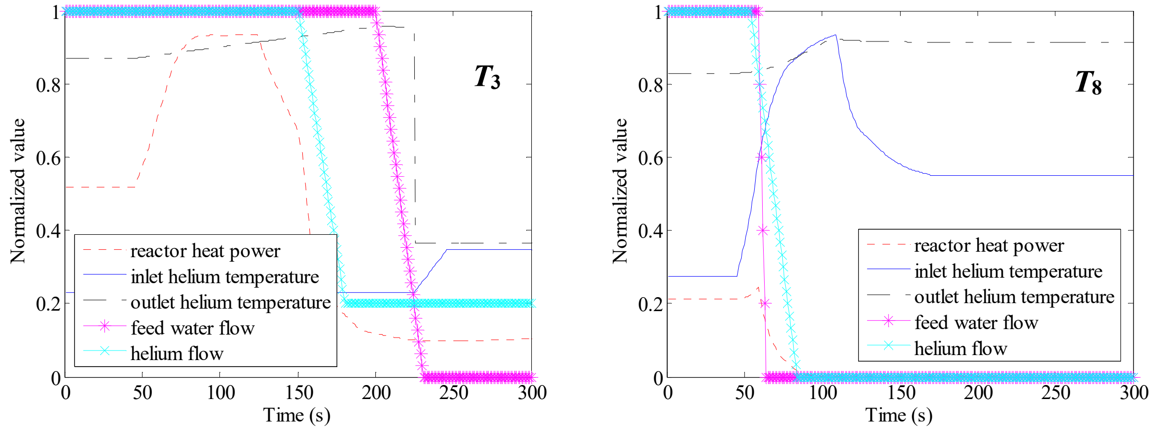

Figure 1. Values of sample data (T_3 and T_8).

4.2 Comparison of the identification results: with and without dimensionality reduction

With the samples of five variables from plant simulator, eigenvalues are calculated and shown in Figure 2. The bigger the eigenvalue is, the more important the related component is. As the eigenvalues decline significantly, we can choose the first several eigenvalues as the main components and then the projection matrix P is obtained by selecting eigenvectors (rows of E^T) with larger eigenvalues. Formula (5) shows that the dimension of $\hat{A}$ can be reduced from $\hat{m}k$ to mk. In this case, the dimension of $\hat{A}$ can be reduced significantly.

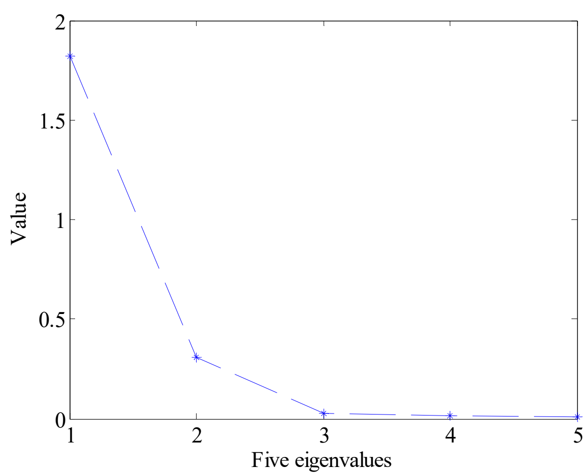

Figure 2. Eigenvalues of covariance matrix C.

The identification results when $rank(P)=1$ are shown in Figure 3, where all results of the identification matched the target classes. That is to say, even $\hat{A}$ is reduced from 1500×11 to 300×11, the proposed method can still identify T_1~ T_{11} that are used to build matrix $\hat{A}$ and the remained T_{12}~ T_{15} that are not contained in $\hat{A}$.

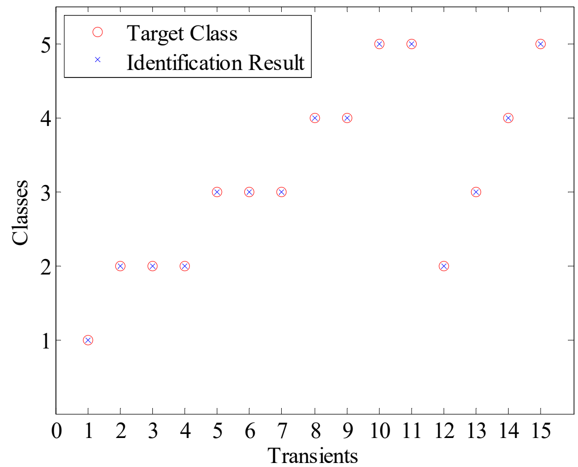

Figure 3. The results of identification (*rank(P)*=1).

5 CONCLUSIONS

Transient identification is very important and beneficial for enhancing the safety and economic efficiency of a nuclear power plant. In this paper, a method, based on PCA, about how to select key features from the state variables is proposed. Then, how to use these extracted features to identify the transients is introduced. The verification results based on the sample data from the simulator of the HTR-PM show that the proposed method can reduce the computation significantly and keep the precision of the identification. Future work will aim at theoretically finding a method to determine the dimension of the projection matrix P for different class of transients.

6 ACKNOWLEDGMENTS

This work has been supported by the Tsinghua University Initiative Scientific Research Program (no. 20151080382 and no. 20151080380).

7 REFERENCES

[1] J. Qianqian, H. Xiaojin, and Z. Liangju, "Operation of Shared Systems via a Common Control System in a Multi-Modular Plant," *Science and Technology of Nuclear Installations*, vol. 2014, pp. 1–9, 2013.

[2] D. Roverso, "Plant diagnostics by transient classification: The ALADDIN approach," *International Journal of Intelligent Systems*, vol. 17, no. 8, pp. 767–790, 2002.

[3] M. G. Na, S. H. Shin, S. M. Lee, D. W. Jung, S. P. Kim, and J. H. Jeong, "Prediction of Major Transient Scenarios for Severe Accidents of Nuclear Power Plants," *IEEE Transactions on Nuclear Science*, vol. 51, no. 2, pp. 313–321, 2015.

[4] Z. Zhang, Z. Wu, D. Wang, Y. Xu, Y. Sun, F. Li, and Y. Dong, "Current status and technical description of Chinese 2×250MWth HTR-PM demonstration plant," *Nuclear Engineering and Design*, vol. 239, no. 7, pp. 1212–1219, Jul. 2009.

[5] B. Jiang, X. Zhu, D. Huang, J. A. Paulson, and R. D. Braatz, "A combined canonical variate analysis and Fisher discriminant analysis (CVA-FDA) approach for fault diagnosis," *Computers and Chemical Engineering*, vol. 77, pp. 1–9, 2015.

[6] J. Ma and J. Jiang, "Applications of fault detection and diagnosis methods in nuclear power plants: A review," *Progress in Nuclear Energy*, vol. 53, no. 3, pp. 255–266, 2011.

[7] Y. Chang, X. Huang, Y. Hao, and C. Li, "Linear Representation and Sparse Solution for Transient Identification in Nuclear Power Plants," *IEEE Transactions on Nuclear Science*, vol. 60, no. 1, pp. 319–327, Feb. 2013.

[8] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed l0 Norm," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, 2009.