Control Engineering Practice 72 (2018) 177–191

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Canonical variable analysis and long short-term memory for fault diagnosis and performance estimation of a centrifugal compressor

Xiaochuan Li [a,*], Fang Duan [b], Panagiotis Loukopoulos [c], Ian Bennett [d], David Mba [a]

[a] Faculty of Technology, De Montfort University, Leicester, LE1 9BH, UK
[b] School of Engineering, London South Bank University, London, SE1 0AA, UK
[c] School of Power and Propulsion, Cranfield University, London, MK430 AL, UK
[d] Rotating Equipment Department, Royal Dutch Shell, Hague, 2501 AN, Netherlands

A R T I C L E   I N F O

Keywords:
Condition monitoring
Long short-term memory
Canonical variable analysis
Fault identification
Performance estimation

A B S T R A C T

Centrifugal compressors are widely used for gas lift, re-injection and transport in the oil and gas industry. Critical compressors that compress flammable gases and operate at high speeds are prioritized on maintenance lists to minimize safety risks and operational downtime hazards. Identifying incipient faults and predicting fault evolution for centrifugal compressors could improve plant safety and efficiency and reduce maintenance and operation costs. This study proposes a dynamic process monitoring method based on canonical variable analysis (CVA) and long short-term memory (LSTM). CVA was used to perform fault detection and identification based on the abnormalities in the canonical state and the residual space. In addition, CVA combined with LSTM was used to estimate the behavior of a system after the occurrence of a fault using data captured from the early stages of deterioration. The approach was evaluated using process data obtained from an operational industrial centrifugal compressor. The results show that the proposed method can effectively detect process abnormalities and perform multi-step-ahead prediction of the system's behavior after the appearance of a fault.

1. Introduction

Modern industrial natural gas processing plants are becoming increasingly complex due to the use of diverse equipment. Because of their complexity, developing an accurate first-principle failure model for such large-scale industrial facilities can be challenging (He, Li, & Bechhoefer, 2012). Thus, existing condition monitoring approaches for industrial processes are typically derived from routinely collected system operating data. Due to the rapid growth and advancement in data acquisition technology, long-term continuous measurements can be taken with sensors mounted on machinery systems. The monitored data are easily stored and analyzed to extract important process condition information.

A number of advanced multivariate statistical techniques have been developed based on condition monitoring data for diagnostic and prognostic health monitoring, such as filtering-based models (Guerra & Kolodziej, 2017), multivariate time-series models (Serdio, Lughofer, Pichler, Buchegger, & Pichler, 2014) and neural networks (Tran, Althobiani, & Ball, 2014). Key challenges in the implementation of these techniques include strongly correlated variables, high-dimensional data, changing operating conditions and inherent system uncertainty (Jiang, Huang, Zhu, Yang, & Braatz, 2015). Recent developments in dimensionality reduction techniques have shown improvements in identifying faults from highly correlated process variables. Conventional dimensionality reduction methods include principal component analysis (PCA) (Harrou, Nounou, Nounou, & Madakyaru, 2013), independent component analysis (ICA) (Hyvärinen, Karhunen, & Oja, 2004) and partial least-squares analysis (PLSA) (Kruger & Dimitriadis, 2008). These basic multivariate methods perform well under the assumption that process variables are time independent. However, this assumption might not hold true for real industrial processes (especially chemical and petrochemical processes) because sensory signals affected by noise and disturbances often show strong correlations between the past and future sampling points (Jiang et al., 2015). Therefore, variants of the standard multivariate approaches (Li & Qin, 2001; Stefatos & Hamza, 2010; Yin, Zhu, Member, & Kaynak, 2015) were developed to solve the time-independency problem, which makes these approaches more suitable for dynamic process monitoring. In addition to approaches derived from PCA, ICA and PLSA, canonical variable analysis (CVA)

* Corresponding author.
E-mail addresses: lix29@lsbu.ac.uk, xiaochuan.li0309@gmail.com (X. Li).

https://doi.org/10.1016/j.conengprac.2017.12.006
Received 5 August 2017; Received in revised form 16 December 2017; Accepted 16 December 2017
0967-0661/

is a multivariate analysis tool. CVA is a subspace method that takes serial correlations between different variables into account and hence is particularly suitable for dynamic process modeling (Stubbs, Zhang, & Morris, 2012). The effectiveness of CVA has been verified by extensive simulation studies (Huang, Cao, Tian, & Deng, 2015; Stubbs et al., 2012) and data captured from experimental test rigs (Cárcel, Cao, & Mba, 2007). However, the effectiveness of CVA in real complex industrial processes has not been fully studied. In the present study, condition monitoring data acquired from an operational industrial centrifugal compressor were used to prove the superior performance of CVA for fault detection and identification in industrial processes.

Once a fault is detected in industrial processes, a prognostic tool is required to predict how the system will behave under faulty operating conditions. Examples of successful applications of different methodologies for performance estimation in the presence of faults are available (Ai et al., 2009; Ai, Zheng, Wang, Jang, & Song, 2010; Zheng et al., 2010). In addition to the abovementioned approaches, CVA is a subspace identification method that can be used to build a dynamic model using measurements of a system's input and output signals. The obtained model can be utilized to predict system performance given expected future input conditions. System inputs used in subspace identification are typically manipulated or controllable variables such as inlet liquid and gas flow valve position. However, the performance of complex industrial systems, such as turbomachines, is not only associated with the system's input signals, which can be manipulated, but is also affected by variations in environmental conditions such as the ambient temperature (Campanari, 2000). The inlet gas temperature of the compressor in this study is a prime example of how environmental conditions can affect a system's performance. Specifically, for the centrifugal compressor, the temperature of the gas to be compressed is largely determined by the ambient temperature when the gas passes through long transmission pipelines to the compressor. As a result, the magnitude of the compressor's inlet gas temperature changes periodically, most commonly every 24 h. To account for the impact of ambient temperature on a system's performance, and thereby allow both the environmental factors and the human interventions to be factored in when predicting the system's future behavior, a time series prediction method is required to forecast the magnitude of the inlet gas temperature based on historical data.

Many data-driven methodologies are available for the prediction of time series, including the widely applied support vector machine (SVM) (Zhang, Wang, & Zhang, 2017), echo state network (ESN) (Chouikhi, Ammar, Rokbani, & Alimi, 2017) and nonlinear auto-regressive moving average (Wootton, Butcher, Kyriacou, Day, & Haycock, 2017) methods. One main challenge of sequence prediction tasks that involve temporal dependencies is handling long-range dependencies (Bengio, Simard, Frasconi, & Member, 1994). Long short-term memory (LSTM) is a powerful learning model that has been extraordinary capable in a wide range of machine learning tasks such as machine remaining useful life prediction (Wu, Yuan, Dong, Lin, & Liu, 2017), visual object recognition (Son, 2017) and speech recognition (Chen & Wang, 2016). LSTM networks use special units in hidden layers that allows inputs to be remembered for long periods; therefore they have great potential in constructing end-to-end systems (Lecun, Bengio, & Hinton, 2015). However, few studies have been conducted to predict sensory signals collected from industrial processes. In this investigation, we explore the ability of LSTM to model the compressor inlet temperature time series. The predicted future inlet gas temperature along with the manipulated system's input signals were fed into a CVA model to perform machine behavior estimation.

The major contributions of this paper are as follows:

- The use of CVA for fault detection using data captured from an operational industrial centrifugal compressor;
- The combination of the canonical state space and the residual space information for fault root-cause analysis;
- The application of LSTM to predict the inlet gas temperature of the compressor in the study;
- The combination of CVA and LSTM for multi-step-ahead prediction of the system's behavior after the occurrence of a fault.

2. Methodology

2.1. CVA for fault detection and identification

CVA is a dimensionality reduction technique used to monitor machine operation by transferring high-dimensional process data into one-dimensional health indicators. Condition monitoring data captured from the system operating under healthy conditions are used to calculate the threshold for normal operating limits. Abnormal operating conditions can be detected when the value of the health indicator exceeds pre-set limits.

The objective of CVA is to maximize the correlation between two sets of variables (Russell, Chiang, & Braatz, 2000). To generate two data matrices from the measured data $y_t \in \mathcal{R}^n$ (n indicates that there are n variables being recorded at each sampling time t), the data were expanded at each sampling time by including a, the number of previous samples, and b, the number of future samples, to construct the past and future sample vectors $y_{a,\,t} \in \mathcal{R}^{na}$ and $y_{b,t} \in \mathcal{R}^{nb}$.

$$y_{a,\,t} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-a} \end{bmatrix} \in \mathcal{R}^{na} \tag{1}$$

$$y_{b,\,t} = \begin{bmatrix} y_{t} \\ y_{t+1} \\ \vdots \\ y_{t+b-1} \end{bmatrix} \in \mathcal{R}^{nb}. \tag{2}$$

To avoid excessive influence of variables with large absolute values, the past and future sample vectors were normalized to zero mean vectors $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$, respectively. Then, the vectors $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$ at different sampling times were rearranged according to Eqs. (3) and (4) to produce the reshaped matrices $\hat{Y}_a$ and $\hat{Y}_b$:

$$\hat{Y}_a = [\hat{y}_{a,t+1}, \hat{y}_{a,t+2}, \ldots, \hat{y}_{a,t+N}] \in \mathcal{R}^{na\times N} \tag{3}$$

$$\hat{Y}_b = [\hat{y}_{b,t+1}, \hat{y}_{b,t+2}, \ldots, \hat{y}_{b,t+N}] \in \mathcal{R}^{nb\times N} \tag{4}$$

where $N = l - a - b + 1$ and l represents the total number of samples for y_t. The Cholesky decomposition was then applied to the past and future matrices $\hat{Y}_a$ and $\hat{Y}_b$ to configure a Hankel matrix $\mathcal{H}$ (Samuel & Cao, 2015). The purpose of using the Cholesky decomposition here is to transfer $\hat{Y}_a$ and $\hat{Y}_b$ into a new correlation matrix with reduced dimensionality such that the subsequent calculations can be conducted in a stable and fast manner. To find the linear combination that maximizes the correlation between the two sets of variables, the truncated Hankel matrix $\mathcal{H}$ is then decomposed using the singular value decomposition (SVD):

$$\mathcal{H} = \sum_{a,a}^{-1/2} \sum_{a,b} \sum_{b,b}^{-1/2} = U \sum V^T \tag{5}$$

where $\Sigma_{a,a}$ and $\Sigma_{b,b}$ are the sample covariance matrices and $\Sigma_{a,b}$ denotes the cross-covariance matrix of $\hat{Y}_a$ and $\hat{Y}_b$.

If the order of the truncated Hankel matrix $\mathcal{H}$ is r, then U, V and $\sum$ have the following form:

$$U = [u_1, u_2, \ldots, u_r] \in \mathcal{R}^{na\times r}$$

$$V = [v_1, v_2, \ldots, v_r] \in \mathcal{R}^{nb\times r}$$

$$\sum = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_r \end{bmatrix} \in \mathcal{R}^{r\times r}.$$