Engineering Applications of Artificial Intelligence 45 (2015) 473–481

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Fault detection with Conditional Gaussian Network

Mohamed Amine Atoui *, Sylvain Verron, Abdessamad Kobi

ISTIA, LARIS EA 7315, L'UNAM, 62, Avenue Notre Dame du Lac, 49000 Angers, France

ARTICLE INFO

Article history:
Received 15 May 2014
Received in revised form
11 June 2015
Accepted 27 July 2015
Available online 28 August 2015

Keywords:
Fault detection
Statistical inference
PCA
Conditional Gaussian Networks
Tennessee Eastman Process
Hot Forming Process

ABSTRACT

The main interest of this paper is to illustrate a new representation of the Principal Component Analysis (PCA) for fault detection under a Conditional Gaussian Network (CGN), a special case of Bayesian networks. PCA and its associated quadratic statistics such as T^2 and SPE are integrated under a sole CGN. The proposed framework projects a new observation into an orthogonal space and gives probabilities on the state of the system. It could do so even when some data in the sample test are missing. This paper also gives the probabilities thresholds to use in order to match quadratic statistics decisions. The proposed network is validated and compared to the standard PCA scheme for fault detection on the Tennessee Eastman Process and the Hot Forming Process.

1. Introduction

Nowadays, systems failures can potentially lead to serious consequences for human, environment or material, and sometimes fixing them could be expensive and even dangerous. Thus, in order to avoid these undesirable situations, it becomes very important and essential for current modern complex systems to early detect any changes in the system nominal operations before they become critical. To do so, several detection methods have been developed and enhanced these last years. These methods can be broadly indexed into two principal approaches, named model-based methods and data-driven methods. Model-based methods are powerful and efficient widely used methods. They are related on the system analytical representation (detailed physical model). However, obtaining this representation for complex, large-scale systems is often not possible or very tricky and requests a lot of time and money. To deal with that, data driven methods have received a significant attention. These methods unlike model-based ones use only measures taken directly from the system (or their transformation) at different times (historical data).

Several data driven methods for faults detection have been proposed (Yin et al., 2012; Ding, 2012; Qin, 2012; Venkatasubramanian et al., 2003; Chiang et al., 2001). Many of them are based on rigorous statistical development of system data and one can mention Subspace aided APproach (SAP), powerful data-driven tools developed to address the problems of building an accurate physical model for complex systems. Partial Least Squares (PLS), Principal Component Analysis (PCA) and their variants (dynamic, non-linear, kernel, and probabilistic) are statistical methods widely used for data reduction and fault detection purpose.

PCA is a well-known and powerful data-driven technique significantly used in application for fault detection but also in many other fields due to its simplicity for model building and efficiency to handle a huge amount of data. In order to identify at any moment if the system is In Control (*IC*) or not (the system is Out of Control *OC*), it is, according to Ding et al. (2010) and Qin (2003), associated to statistics with quadratic forms. These statistics are not only associated to PCA but also to many others data driven and model-based methods. Among these statistics, two well-known and used statistics are the T^2 and SPE (Squared Prediction Error) statistics. These two are generally combined to complement each other and thus enhance the fault sensitivity.

Meanwhile, in the last decades, Bayesian networks (BN) have been also proposed for fault detection (Yu and Rashid, 2013; Verron et al., 2010a; Huang, 2008; Roychoudhury et al., 2006; Schwall and Gerdes, 2002; Lerner et al., 2000). BN's are powerful tools designed by experts and/or learned from data. They offer a Probabilistic/statistical framework that able to integrate information from different sources which may be of interest for fault detection. Indeed, the use and the fusion of all the information available on the system (as causal influences (e.g. graphical representations of variables dependencies), probabilistic fault detection decisions, maintainability information, components reliability and so on) could enhance and provide better decisions. On this perspective, we propose to use a BN in order to model PCA fault detection techniques.

* Corresponding author.
E-mail address: mohamed.atoui@univ-angers.fr (M.A. Atoui).

http://dx.doi.org/10.1016/j.engappai.2015.07.020
0952-1976/

Nomenclature

0	zero or a vector of zeros, depending on the context
$B_i, \hat{B}_i, \tilde{B}_i$	respectively rows of the eigenvectors matrix P, $\hat{P}$, $\tilde{P}$
c_Δ	coefficient of variability
$\mathbf{e}$	natural exponential function
E	set of an BN arcs
F	Fisher distribution
g_α	the normal deviate corresponding to the upper $1-\alpha$ percentile
I	identity matrix
$\mathcal{N}(\mu;\Sigma)$	Gaussian (normal) probability density function (pdf) with μ-means and covariance matrix Σ
N	number of samples
$P, A, \hat{P}, \hat{A}, \tilde{P}, \tilde{A}$	eigenvectors matrices
p	a probability measure: a probability distribution or a probability density function. Its meaning will be clear from the context
V	set of an BN nodes
$\mathbf{x}, \mathbf{x}^+, \mathbf{x}^-, \mathbf{y}, \mathbf{t}$	multivariate variables
X	normalized set of samples of $\mathbf{x}$
$\hat{X}, \tilde{X}$	respectively the systematic and noise part of X
$Z, \hat{Z}, \tilde{Z}$	spaces generated by PCA
α	value of the error of first kind
ϵ	error model
Λ	set of non-negative real eigenvalues
ω	ratio of $p(OC)$ to $p(IC)$
CL_Δ	control limit of the quadratic statistic Δ
$\theta_1, \theta_2, \theta_3, h_0$	parameters used for the calculation of CL_{SPE}
$\zeta_\Delta^{IC}, \zeta_\Delta^{OC}$	probabilistic control limits given the quadratic statistic Δ and the states : IC, OC
$Pa(\mathbf{x})$	set of parent nodes of $\mathbf{x}$
$A \backslash \{B\}$	A except B
$\mathbb{E}[\mathbf{x}], \mathrm{cov}[\mathbf{x}]$	respectively the expected value and the covariance of the variable $\mathbf{x}$

Another important challenge is to handle on-line missing observations. The most used approaches are based on the imputation methods, which try to complete the missing values. However, these methods are time consuming and depend strongly on the missing rate of the original sample. The proposed network, unlike most of the proposed Bayesian networks for fault detection, is able to respect a false alarm rate, model PCA fault detection scheme and handle automatically missing observation without delay or imputation. The main interests of this paper can be described in few points : (1) a generalized form of the quadratic statistics (e.g. T^2, SPE) under a probabilistic tool, (2) a probabilistic framework for fault detection purpose, managing both PCA (systematic and residual subspaces) and statistics under a single BN using discrete and Gaussian nodes, and (3) probabilities about the system state could be provided, even when data on line are missing (a non-imputation method to handle unobserved variable s).

The remainder of this paper is structured as follows. In Section 2 a brief description of some definitions and tools needed to develop our proposals is given, Section 3 describes and introduces the development of PCA under CGNs for fault detection purpose. This is followed by a comparison between our proposal and the standard PCA, two cases studies are given. Finally, conclusions and outlooks are outlined in the last section.

2. Tools

2.1. Bayesian Networks

2.1.1. Definition

A Bayesian Network (BN) (Jensen and Nielsen, 2007) is a probabilistic graphical model. It is associated and consists of the following:

- a directed acyclic graph $\mathbf{G}$, $\mathbf{G}=(V,E)$, where V is the vertexes set of $\mathbf{G}$ (nodes), and E is the edges set of $\mathbf{G}$ (arcs),
- a finite probabilistic space $(\Omega, \mathbb{Z}, p)$, with Ω a non-empty space, $\mathbb{Z}$ a collection of the subspaces of Ω and, p a probability measure on $\mathbb{Z}$ with $p(\Omega)=1$,
- a set of random variables $\mathbf{x}=\mathbf{x}_1,\dots,\mathbf{x}_m$ associated with the vertexes of the graph $\mathbf{G}$ and defined on $(\Omega,\mathbb{Z},p)$, such that:

$$p(\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_m)=\prod_{i=1}^{m} p(\mathbf{x}_i \mid Pa(\mathbf{x}_i)) \tag{1}$$

  where $Pa(\mathbf{x}_i)$ is the set of the parent nodes of $\mathbf{x}_i$ in $\mathbf{G}$,
- a conditional probability table (CPT) associate to each node, given its parents, describing probabilistic dependencies between variables,
- calculations (e.g. based on Bayes rule) named inference, used given the availability of new information (evidence) about one or several $\mathbf{G}$ nodes, to update the network (e.g. to give the posterior probabilities).

2.1.2. Conditional Gaussian Networks

A particular form of Bayesian networks is the Conditional Gaussian Network (CGN). Each node in the network represents a random variable that may be discrete or Gaussian (univariate/multivariate). However, following Lauritzen and Jensen (2001), Lauritzen (1992), for the availability of exact computation (inference) discrete nodes are not allowed to have continuous parents, they have only discrete parents. Thus, each Gaussian node, given its Gaussian parents follows a Gaussian linear regression model (linear combination of its continuous parents observations), with parameters depending on its discrete parents. In this paper, we restrict our attention to two kinds of Gaussian nodes.

First, the linear Gaussian node, a Gaussian node $\mathbf{y}$ with only Gaussian parents $\Phi_1,\dots,\Phi_d$. Its conditional distribution is given by

$$p(\mathbf{y} \mid \Phi_1=\phi_1,\dots,\Phi_d=\phi_d)=\mathcal{N}(\mu_{\mathbf{y}}+W_1\phi_1+\dots+W_d\phi_d;\ \Sigma_{\mathbf{y}}) \tag{2}$$

where $\mu_{\mathbf{y}}$ is a parameter governing the mean of $\mathbf{y}$, $\Sigma_{\mathbf{y}}$ is the covariance matrix of $\mathbf{y}$, $W_1,\dots,W_d$ are the regression coefficients. Note that, the joint distribution p $(\mathbf{y}, Pa(\mathbf{y}))$ is also Gaussian. If $\Sigma_{\mathbf{y}}$ is null then (2) represents a deterministic linear relationship between $\mathbf{y}$ and its parents.

The second node, the conditional linear Gaussian node without Gaussian parents, a Gaussian node $\mathbf{y}$ with only discrete parents $Pa(\mathbf{y})=(\Theta_1,\dots,\Theta_d)$. It is linear Gaussian for each value $k_{Pa(\mathbf{y})}$ of its parents $Pa(\mathbf{y})$. Its conditional distribution could be written as below:

$$p(\mathbf{y} \mid Pa(\mathbf{y})=k_{Pa(\mathbf{y})})=\mathcal{N}(\mu_{k_{Pa(\mathbf{y})}};\Sigma_{k_{Pa(\mathbf{y})}}),\quad k_{Pa(\mathbf{y})}\in K_{Pa(\mathbf{y})} \tag{3}$$

where $\mu_{k_{Pa(\mathbf{y})}}$ and $\Sigma_{k_{Pa(\mathbf{y})}}$ are respectively the mean and the covariance matrix of $\mathbf{y}$ given the values $k_{Pa(\mathbf{y})}$ of its parents. $K_{Pa(\mathbf{y})}$ represent the different values that the parents of $\mathbf{y}$ can take.

2.1.3. Discriminant analysis and CGN

Many Conditional Gaussian Networks can be used to solve discrimination problems between different data classes. Their nodes set V always include a discrete node indexing the different