Computers and Chemical Engineering 60 (2014) 260–276

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Optimal variable selection for effective statistical process monitoring

Kaushik Ghosh[a,1], Manojkumar Ramteke[b,2], Rajagopalan Srinivasan[a,b,*,3]

[a] Department of Chemical and Biomolecular Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore
[b] Process Sciences and Modeling, Institute of Chemical and Engineering Sciences, 1 Pesek Road, Jurong Island, Singapore 627833, Singapore

ARTICLE INFO

Article history:
Received 2 November 2012
Received in revised form
17 September 2013
Accepted 18 September 2013
Available online 25 September 2013

Keywords:
Fault detection
Optimization
Process control
Safety
Systems engineering
Tennessee Eastman Process

ABSTRACT

In a typical large-scale chemical process, hundreds of variables are measured. Since statistical process monitoring techniques typically involve dimensionality reduction, all measured variables are often provided as input without weeding out variables. Here, we demonstrate that incorporating measured variables that do not provide any additional information about faults degrades monitoring performance. We propose a stochastic optimization-based method to identify an optimal subset of measured variables for process monitoring. The benefits of the reduced monitoring model in terms of improved false alarm rate, missed detection rate, and detection delay is demonstrated through PCA based monitoring of the benchmark Tennessee Eastman Challenge problem.

1. Introduction

Process monitoring plays an important role in the safe operation of chemical processes. The task of monitoring involves detection of abnormal process behavior. For successful monitoring, it is important to detect faults as early as possible. Early detection of fault not only enables prevention of major accidents but also reduces the maintenance downtime, and is instrumental in maintaining plant throughput. Process monitoring techniques can be broadly classified into: model-based and data-based methods (Chiang, Russell, & Braatz, 2001; Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003a,c; Venkatasubramanian, Rengaswamy, & Kavuri, 2003b). Model-based techniques require a comprehensive mathematical model which is often difficult or costly to obtain for complex chemical processes. Data-driven techniques have therefore become attractive not only because process data are readily available for use in many process systems, but also because the techniques do not need any rigorous models.

Among data-driven techniques, a popular category is multivariate statistical process monitoring (MSPM) methods such asprincipal component analysis (Kresta, MacGregor, & Marlin, 1991), partial least square (Piovoso & Kosanovich, 1994), fisher discriminant analysis (Chiang, Russell, & Braatz, 2000), and Correspondence Analysis (Detroja, Gudi, & Patwardhan, 2007). Apart from these, various pattern classification based methods such as artificial neural network (Srinivasan, Wang, Ho, & Lim, 2005a,b), Support vector machine (Mahadevan & Shah, 2009), self organizing map (Ng & Srinivasan, 2008a,b), and *k*-nearest neighbor (He & Wang, 2007) have also been used extensively. Artificial immune system (AIS) is a relatively new data-driven methodology wherein the principles and processes of the natural immune system are abstracted and applied in solving real-world problems. Recently, AIS has attracted much attention for process monitoring and fault diagnosis (Dai & Zhao, 2011; Ghosh & Srinivasan, 2011; Zhao et al., 2013). Monitoring performance of the data-driven techniques depends largely on the quality and quantity of the data used to build the monitoring model. Modern chemical plants contain a large number of mass- and heat-integrated unit operations and thousands of process variables are measured regularly. Not all the measured variables are however equally important for process monitoring. Usually, only a small subset of the recorded variables carries essential information about the faults that can affect the process and is hence more useful for developing a monitoring model. The rest of the paper is organized as follows: in Section 2, we provide a brief literature review on feature selection, a related topic widely studied in the pattern recognition literature. In Section 3, the effect of variable selection on the performance of PCA based process monitoring method is illustrated. Based on those insights, a GA based variable selection scheme is presented in Section 4. In Section 5, the efficacy of the proposed scheme is illustrated

* Corresponding author. Current address: Indian Institute of Technology Gandhinagar, India.
E-mail address: raj@iitgn.ac.in (R. Srinivasan).
[1] Current address: ABB Corporate Research Center, Bangalore 560048, India.
[2] Current address: Department of Paper Technology, Indian Institute of Technology, Roorkee 247667, India.
[3] Current address: Indian Institute of Technology, Gandhinagar 382424, India.

0098-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2013.09.014

through the benchmark Tennessee Eastman process challenge problem.

2. Feature selection

In classical pattern recognition, patterns are generally represented as a vector of feature values. The selection of a subset of features can have a considerable impact on the resulting classification performance (Jain & Zongker, 1997; Raymer, Punch, Goodman, Kuhn, & Jain, 2000). It has been shown experimentally that irrelevant and noisy features unnecessarily increase the complexity of the classification problem and can degrade performance (Na, Sim, & Park, 2002; Zio, Balardi, & Pedroni, 2006). As a result, feature selection methods have become an important pre-processing step in automated pattern recognition (Biem, Katagiri, & Juang, 1997; Jain & Zongker, 1997; Lee & Landgrebe, 1997; Raymer et al., 2000), exploratory data analysis (Mao & Jain, 1995), and data mining (Raymer et al., 2000).

Feature selection techniques study how to identify and select informative (discriminative) features for building classifiers that can interpret data better. Feature selection can reduce the computational cost by reducing data dimensionality, increase classification speed and accuracy, and improve the comprehensibility of the classifier models by eliminating redundant and irrelevant (probable noise) features. It has to be noted that feature selection is different from feature extraction (or dimensionality reduction) which creates new features by combining the original features. In general terms, feature extraction refers to the task of finding a mapping that reduces the n-dimensional data being transformed onto a m-dimensional space, where $m < n$ (Zio et al., 2006). All n original features are used for obtaining the m-dimensional transformed data. On the other hand, feature selection maintains the original meaning of the selected features. Thus, feature selection is a special case of feature extraction where by $(n - m)$ irrelevant features are discarded and only the m most informative ones are retained. Feature selection and extraction have many functions in common. Both can be used to project data onto lower dimensional space for subsequent visualization, clustering, and other exploratory data analysis (Raymer et al., 2000).

Feature selection algorithms can be divided into two categories: filter-based and wrapper-based methods (Tan, Fu, Zhang, & Bourgeois, 2008; Zio et al., 2006). Filter-based methods utilize the intrinsic properties of the data to select subsets of features as a preprocessing step, independently of the chosen classifier. Basically, they discard irrelevant and/or redundant features before the construction of the classifier. This is independent of the specific learning algorithm used in the classification. Features are assessed by their relevance or discriminant powers with regard to the targeted classes. Feature ranking approaches score or rank features by certain criterion and use the rankings as the basis for selection. These are particularly attractive because of their simplicity, scalability, and good empirical success (Tan et al., 2008). Computationally, feature ranking is efficient since it requires only the computation of a score for each feature and sorting them. Based on the scores, subsets of significant features can be selected to build a classifier. Some feature selection methods use criteria based on statistics, such as χ^2 statistics (Liu & Setiono, 1995), T statistics (Liu, Li, & Wong, 2002), F statistics (Peng, Long, & Ding, 2005), signal-to-noise statistic (Golub et al., 1999), Fisher criteria (Furey, Cristianini, Bednarski, & Schummer, 2000), information gain (Liu, 2004), mutual information (Guyon & Elisseeff, 2003; Peng et al., 2005), and entropy-based measures (Dash & Liu, 1999; Liu, Krishnan, & Mondry, 2005). A common drawback of many feature ranking methods is that they implicitly assume that features are orthogonal (i.e., uncorrelated) to each other and assess features in isolation. Features are ranked on the basis of their individual predictive capabilities. Mutual information such as redundancy or complementariness among features is ignored. In fact, top-ranked features might be strongly correlated so that using two or more of them may not provide any added benefit. This is common in chemical processes where variables are connected through mass and energy integration. In addition, a feature which is insignificant according to some feature ranking or selection measurement can provide a significant performance improvement when grouped with other features (Srinivasan & Qian, 2007). These issues of redundancy and multivariate prediction limit the application of feature ranking algorithms in large-scale chemical processes.

Wrapper-based methods select a subset of the features according to the classification performance of the chosen classifier, and hence consider the mutual dependency among features (Hamdani, Won, Alimi, & Karray, 2011). They allow simultaneous feature selection and classifier training to produce the optimal combination of features and classifiers for a particular classification problem. GAs (Goldberg, 1989) has been widely used in wrapper-based feature selection to efficiently explore the combinatorial solutions space (Li et al., 2011). Oh, Lee, and Moon (2004); Tan et al. (2008); Zhu and Guan (2004) demonstrated that the solutions identified by GA were more efficient in terms of achieving maximum classification accuracy than other classical feature selection methods.

Wrapper-based approaches generally outperform filter-based methods in terms of prediction accuracy since the former ensure the selection of features more suitable for the classification algorithm used, whereas the latter completely ignore the effects of the selected feature subspace on the performance of the classifier. Filter-based approaches are generally computationally more efficient since they avoid the additional steps of classifier learning and evaluating the performance of the learned classifier which are computationally expensive. However, since the feature selection step is usually performed in the offline phase prior to classifier training, the increase in computational cost is not a critical issue. Improved prediction accuracy clearly outweighs computational consideration. Hence, in this paper, we use a wrapper-based approach.

Although feature selection is widely used as a first step in pattern classification and data mining applications it has not been considered in multivariate statistical process monitoring (MSPM) with the notable exception of Verron, Tiplica, and Kobi (2008). Design of sensor network and optimal sensor placement, which has received much attention in process systems engineering literature, is closely related conceptually to feature selection. There, the main objective is to select the locations for sensors so as to maximize estimation accuracy (Kretsovalis & Mah, 1987), minimize capital cost (Bagajewicz, 1997), or maximize some fault detection criterion (Bhushan & Rengaswamy, 2002a,b; Musulin, Bagajewic, Nougues, & Puigjaner, 2004). MSPM method based on PCA is a dimensionality reduction technique (i.e., feature reduction). Therefore, at the outset, variable selection (or feature selection) may appear to be redundant. However, as shown next, variable selection can have a significant impact on the monitoring performance of PCA.

3. Effect of variable selection on monitoring performance

Consider an artificial process with three variables (x_1, x_2 and x_3) that follow standard normal distribution when the process is normal.

$$\begin{aligned} x_{1,N} &\in N(0, 1) \\ x_{2,N} &\in N(0, 1) \\ x_{3,N} &\in N(0, 1) \end{aligned} \tag{1}$$