Journal of Process Control xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Process Control

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

Research Paper

Large-scale plant-wide process modeling and hierarchical monitoring: A distributed Bayesian network approach

Jinlin Zhu [a], Zhiqiang Ge [a,*], Zhihuan Song [a], Le Zhou [b], Guangjie Chen [a]

[a] State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, PR China
[b] School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou, 310027, PR China

ARTICLE INFO

Article history:
Received 11 June 2016
Received in revised form 3 August 2017
Accepted 30 August 2017
Available online xxx

Keywords:
Plant-wide process monitoring
Distributed modeling
Hierarchical monitoring
Bayesian network
Bayesian decision fusion

ABSTRACT

In this work, a systematic distributed Bayesian network approach is proposed for modeling and monitoring large-scale plant-wide processes. First, to deal with the large-scale process modeling issue, the entire plant-wide process is decomposed into blocks and Bayesian networks are constructed for different blocks. Subsequently, distributed Bayesian network blocks are fused into a global Bayesian network with a proper designed algorithm. For fault detection, a missing data approach is proposed for state estimation, based on which the T^2 and Q statistics are constructed. Finally, a Bayesian decision fusion mechanism is established for hierarchical monitoring of variables, unit blocks and the global industrial plant. For fault isolation, a Bayesian contribution index is further developed and the corresponding isolation scheme is proposed. Simulation results on the plant-wide Tennessee Eastman process show that the distributed Bayesian network approach can be feasible for modeling large-scale process. Furthermore, the proposed hierarchical monitoring scheme provides informative multi-level reference results for further diagnosis and isolation.

1. Introduction

Plant-wide process modeling and monitoring have become popular over the past few decades due to the ever-increasing complexity of modern industrial systems [1–4]. The plant-wide process usually refers to a process with multiple operating units, workshops and even plants [5]. Therefore, the whole plant-wide process is typically of large-scale in size and also of high complexity in variable correlations [6]. Compared to traditional industrial systems, modeling and monitoring for plant-wide processes should be more challenging [7].

In order to ensure normal operation of the plant-wide process, safety monitoring mechanisms should be employed. Practically, for process with in-depth analysis of process knowledge, one can resort to model-based first principle methods to build observers for estimating process states and identifying those malfunctioned sensors [8]. However, mechanistic analysis and kinematic equation inductions for a large system could be laborious and even unavailable which in turns limit the application of model-based methods only for small-scale systems [9]. For large-scale cases, data-driven methods require little system knowledge for process modeling can be regarded as good alternatives [10–17]. It is noticed that unlike traditional small-scale processes with centralized monitoring techniques, multiple blocks are usually divided for plant-wide processes according to process system sub-sections and then multivariate statistical process monitoring models could be applied. Several benefits of multiblock modeling techniques can be found. First, it can reduce the modeling complexity due to the fact that the global large system can be commonly decomposed by several smaller local sections which are connected by only a few streams [18]. Second, multiblock methods can better describe local operating systems since an explicit block of multivariate statistical model can be defined on each section. Moreover, by dividing blocks, one can detect faulty events in a decentralized way and can also locate or isolate the most responsible faulty sections as well as faulty variables by checking control charts and contribution charts respectively. Under this framework, the multiblock principal component analysis (MBPCA) and multiblock partial least square (MBPLS) modeling and monitoring methods have been developed [19]. Qin analyzed both multi-block methods and introduced the decentralized monitoring/diagnosis framework [20]. The concurrent decentralized monitoring diagram has also been proposed recently to localize input and output relevant faulty events [21]. It should be noticed that the aforementioned multiblock methods

* Corresponding author.
E-mail address: gezhiqiang@zju.edu.cn (Z. Ge).

http://dx.doi.org/10.1016/j.jprocont.2017.08.011

directly make use of process system sub-sections. Recently, some mechanisms have also attempted to realize automatic block dividing. For example, Ge and Song proposed a distributed PCA scheme that divides blocks according to the variations of multiple directions of principal components [7]. Jiang and Yan conduct block division by calculating mutual information for all variables to test the correlations based on the assumption that process variables within the same block should show strong correlations, then a specific PCA is built on each block for monitoring [22]. Recently, Jiang and Huang have proposed the heuristic-based process decomposition and Bayesian diagnostic schemes so as to derive performance-driven monitoring results [23]. However, most of the current data-driven methods are built based on the large number of measurement and manipulation variables collected accordingly by those hardware sensors widely installed on each operating unit, regardless of some obvious or latent process knowledge. As a typical model, PCA simply captures the quantitative relations among orthogonal projections on large variances and builds the control limits for normal operating regions. However, some qualitative process knowledge such as the causality among variables has been neglected.

Generally speaking, before designing proper monitoring mechanisms for plant-wide processes, one can often obtain some global characteristics such as process equipment as well as the forward/feedback interactions among them based on the system flow diagram. For instance, if one defines two blocks by operating units say a reactor and a separator, then the causality among variables within and between these two blocks should also be important information as a part of process knowledge. Otherwise, if the reactor and separator show weak connections in process flowchart, these two blocks can be modeled independently. Therefore, all these qualitative process knowledge should be treated rationally for a reasonable quantitative modeling procedure, especially for a large-scale plant-wide process where the flow diagram is often intuitive and informative.

To make better use of the process knowledge, the Bayesian network approach has become popular by describing variable relations from the perspective of both quantitative and qualitative scales using causality structure and conditional probabilistic assignation respectively [24]. Verron et al. proposed a static Bayesian network approach for fault detection and isolation [25]. The static network structure is learned from data directly with a specific intelligent algorithm called PC algorithm. For monitoring, the T^2 statistic is decomposed according to the causal map of the Bayesian network so that the faulty variables can be diagnosed and isolated only by those most probable responsible variables. Yu and Rashid proposed a dynamic Bayesian network approach for process monitoring [24]. In this work, the Bayesian network structure is constructed totally based on the process knowledge, and then probabilistic indexes, namely abnormal likelihood index and dynamic Bayesian indexes are designed for fault detection, propagation identification and root cause analysis. On the basis of this work, Mori et al. proposed a sampling-based approach for structure building also with the analysis of process knowledge [26]. For monitoring, the authors use a sampling-based method for identifying the most probable propagation path. It should be noticed that the obtained probabilistic model is essentially a structural equation model instead of a Bayesian network since the acyclic assumption for network structure is not satisfied. Nevertheless, the probabilistic network approach has proved to be successful by all these works for industrial applications.

However, for plant-wide modeling and monitoring, there are still some issues that have not been well considered. First, the distributed Bayesian network modeling should be reconsidered for large-scale application. In other words, one should not only build Bayesian networks for sub-sections but also should consider how to elegantly fuse all distributed models into a global Bayesian network for the convenience of Bayesian learning and inference. Second, the monitoring statistics should be constructed for abnormal event detection. In the previous works, the monitoring statistics are either established with just T^2 statistics alone or with some likelihood-based probabilistic indexes. However, it could be more reasonable if one can monitor both of the variable space (with T^2 statistics) and the residual space (with Q statistics) like those widely applied in traditional methods. Besides, the distributed monitoring results need to be explicitly fused as a global index from a suitable way for indicating the entire process operating status. For the plant-wide process, due to the large scale with hundreds of variables, simply monitoring from the variable level could be laborious for fault analysis and process status visualization. Alternatively, it could be more desirable to monitor the large-scale process from variables, blocks and the global plant in a hierarchical way.

In the present paper, a systematic distributed Bayesian network approach is proposed for effective distributed modeling and hierarchical monitoring of large-scale plant-wide processes. First, the entire process is decomposed into several blocks, based on which distributed Bayesian network models are developed in a parallel manner for all blocks. During this step, the process knowledge is fully utilized for distributed modeling such as process decomposition. Afterwards, a fusion algorithm is established for combining all distributed networks into a single Bayesian network. The global Bayesian network is composed of the qualitative information which is denoted by causality relationships of all operating variables and also the quantitative information which is represented with the explicit conditional likelihood distributions assigned for all network variables. For monitoring, a missing variable approach is proposed for Bayesian inference of all network variables and then both T^2 and Q statistics are designed for implicated network variables. Finally, all corresponding statistics are fused with a Bayesian strategy to form the global plant-wide Bayesian index. As a further analysis, the faulty variable and faulty block contribution indexes are also proposed for fault isolation. From this way, one can systematically model plant-wide process in a distributed way and also monitor the system operating behaviors of the large-scale process in the hierarchical manner from the variable level, distributed block level and the global plant-wide level.

The rest of this paper is organized as follows. First, some preliminaries on Bayesian networks are given. Then, the systematic distributed modeling approach for Bayesian networks is proposed. In Section 4, the monitoring statistics are developed and the Bayesian fuse mechanism is designed hierarchically for plant-wide monitoring. For industrial application, the proposed method is evaluated on the Tennessee Eastman benchmark process. In the last section, conclusions and some outlooks are summarized.

2. Preliminaries

2.1. Bayesian network definition

Generally, a Bayesian network structure can be defined by a directed acyclic graph (DAG) G with a node set V and an edge set E, which is usually denoted by $G = \left\{V, E\right\}$. Each node in DAG is a measurement variable written as $V = \left\{X_1, X_2, \ldots, X_I\right\}$ where I is the number of variables [27]. Each directed edge connects two nodes from a parent node (i.e. cause of an event) to a children node (i.e. effect of an event) and no cycles are involved. Commonly, variables with no parents are called root nodes while variables with no child nodes are called leaf nodes. In addition to the qualitative relations represented by DAG, the quantitative details are described according to the statistical probability distributions defined which are further simplified by statistical dependencies among the variable nodes. An important property for the network is the conditional

Please cite this article in press as: J. Zhu, et al., Large-scale plant-wide process modeling and hierarchical monitoring: A distributed Bayesian network approach, J. Process Control (2017), http://dx.doi.org/10.1016/j.jprocont.2017.08.011