G Model
JJPC-2197; No. of Pages 11

ARTICLE IN PRESS

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

Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes

Zhiqiang Ge

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, PR China

ARTICLE INFO

Article history:
Received 21 October 2016
Received in revised form 15 July 2017
Accepted 30 August 2017
Available online xxx

Keywords:
Plant-wide processes
Key performance index
Predictive modeling
Diagnosis
Distributed data modeling

ABSTRACT

In this work, a distributed predictive modeling framework is proposed for prediction and diagnosis of key performance indices in plant-wide processes. With block division of the plant-wide process, key data information can be extracted more efficiently, based on which the predictive model can then be developed for regression of the key performance indices. In order to determine the root causes of performance degradation for the key performance index, a diagnostic scheme is developed among this framework. First, the critical blocks are identified through definition of the block contribution in the diagnostic model. The contribution of each process variable is then evaluated inside each critical block, based on which the root causes of performance degradation can be successfully located. An example of the distributed modeling method is realized by using the basic Principal Component Analysis and Gaussian Process Regression models, with a detailed case study on the TE benchmark process.

1. Introduction

Due to the wide use of the distributed control system (DCS), a large amount of data has been recorded in the process industry [1–3]. While those ordinary process variables such as temperature, flow, and pressure have been routinely recorded, some key performance indices may not be able to be measured online. For example, measuring the melt index in the polypropylene production process is very difficult, the value can only be obtained through laboratory analysis; in a typical batch process, the quality index of the final product cannot be measured until the end of the batch. Therefore, the quality performance index can only be predicted through the whole batch duration, which is important for quality control in the batch process. Furthermore, for construction of an effective quality control system, any degradation of those key performance indices should be diagnosed and the root causes need to be located. As a matter of fact, prediction and diagnosis of key/quality performance indices has recently become a hot research topic in the process systems engineering area [4–8].

Basically, predictive modeling methods can be partitioned into two categories, termed as model-based methods and data-based methods. Compared to the model-based method which has significant reliance on process mechanism and knowledge, the data-based method has no such restriction thus can be used in lots of complex industrial processes. To date, various data-based predictive modeling methods have been developed, such as principal component regression (PCR), partial least squares (PLS), artificial neural network (ANN), support vector regression (SVR), etc [9–16].

This paper is focused on the plant-wide industrial process which may consist of several different operation units, workshops or even manufacturing plants. While monitoring for plant-wide processes has become a quite hot research spot in recent years [17–26], prediction and diagnosis for key performance indices in those processes have rarely been reported. Compared to the ordinary process, prediction and diagnosis of key performance indices in the plant-wide process is much more difficult. For example, different parts of the plant-wide process may have simultaneous and cumulative effects on the final product quality; the contributions of different parts of the process on estimating the final product quality may vary from each other; predictive modeling with a large number of process variable may also cause a computational problem, especially for those methods which have already involved high computational burdens.

In the present paper, a distributed predictive modeling framework is proposed for prediction and diagnosis of key performance indices in plant-wide processes. The whole plant-wide process is first divided into several different parts, which usually correspond to different units or sections of the process. Second, individual models are developed in each part of the process, based on which

E-mail address: gezhiqiang@zju.edu.cn

http://dx.doi.org/10.1016/j.jprocont.2017.08.010
0959-1524/© 2017 Published by Elsevier Ltd.

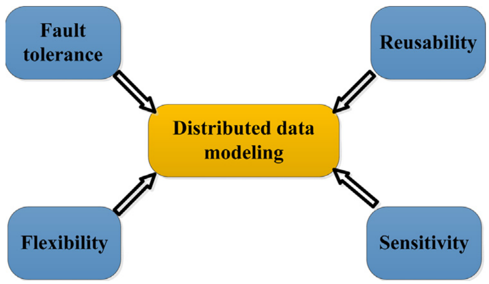

Fig. 1. Advantages of the distributed modeling framework.

the important data information can be efficiently extracted, while simultaneously the dimensionality of the data could be greatly reduced. As a result, the computational burden of the predictive modeling process can be significantly reduced. With the extracted information from different parts of the process, an overall predictive model can then be constructed to connect them with key performance indices of the process.

Another advantage of the distributed modeling approach is due to the facilitated diagnosis of key performance indices. In the plant-wide process, identification of those parts which are critical to the key performance index is particularly interested. As long as the critical-to-performance parts have been identified, the key performance index of the process can be controlled more accurately and the improvement can be made more efficiently. Furthermore, after the critical parts have been identified, the key performance index related process variables in those corresponding parts can be further located based on the individual models.

The rest of this paper is organized as follows. In Section 2, the main idea of the distributed predictive modeling framework is illustrated, including a detailed example based on PCA and GPR models, discussions and remarks on related issues regarding the distributed predictive modeling framework. Then, a case study of a typical benchmark process is carried out in Section 3, in order to demonstrate the feasibility of the distributed predictive modeling method. Finally, conclusions are made and some outlooks are highlighted for future research on this topic.

2. Methodology

In this section, the framework of the distributed predictive modeling method is demonstrated, including the idea of the general framework, an example based on PCA and GPR models, and discussions and some remarks on related issues among this modeling framework.

2.1. Distributed predictive modeling framework

As have been mentioned in the introduction part, there are some related methods that have already been developed for plant-wide process monitoring. Among them, a new distributed modeling framework has recently been proposed, which provided a general framework for distributed monitoring for plant-wide processes. The main advantages of the distributed modeling framework are summarized in Fig. 1: the fault tolerance ability of the process can be improved under the distributed modeling framework; it is more flexible than the central modeling framework; sensitivity analysis for the model and performance evaluation will be significantly improved under the distributed modeling framework; and the computational complexity of the modeling procedure can be significantly reduced [27].

With the incorporation of the key performance indices, a distributed predictive modeling framework can be formulated for plant-wide processes. The detailed flowchart for development of a distributed predictive data model is provided in Fig. 2. In the first step, the whole plant-wide process needs to be divided into different parts/blocks. This can be done through implementation of effective process knowledge, engineering experiences, or some expert systems. Alternatively, those statistical analysis approaches which are totally data-driven can also be possibly used for block division. However, our experience is that it could be better to use the process knowledge if there is any useful information about it. In contrast, the data-driven method can be used as a complementary tool, to revise any inappropriate result, or to make a further block division if necessary. In the second step, training dataset is extracted from the historical databased, with the help of sample selection and variable selection methods. Actually, this is a very important step for construction of the distributed predictive modeling framework, because both samples and variables play important roles in data modeling. Without selections of appropriate samples and variables in each block of the process, the performance of the data model cannot be guaranteed, the key information cannot be well extracted, and thus it could influence the performance of the predictive model. In order to do effective sample and variables selections, various methods can be applied, such as abnormal detection, time series analysis, correlation analysis, quality related analysis, and so on.

After the process has been divided into different blocks and the modeling datasets have already prepared inside each block, the next step is to do block modeling and information extraction. This is related to the model selection problem. Generally, the type of the data model should be selected according to the characteristic of the modeling data inside each block. The aim of this step is to extract key information from each block, and simultaneously reduce the dimensionality of the variables. Therefore, multivariate statistical modeling and dimensionality reduction approaches are particularly useful, such as PCA, PLS, ICA and so on. After the key information has been extracted from all of the blocks, they can be integrated together for predictive modeling in the next step.

The key step of the distributed predictive modeling framework is the data regression modeling between the integrated extracted information from multiple blocks and the information of key performance indices. In this step, the structure of the regression model needs to be carefully selected. If there is high nonlinear relationships between those two datasets, a nonlinear regression modeling method should be selected; if the dynamical nature of the data is significant, a dynamical model may be more useful; if a part of the process variables are non-Gaussian distributed, a non-Gaussian regression model could be applied; etc. Furthermore, the performance of the regression model should be evaluated. If there is any change of the modeling environment or the performance of the regression model degrades, the model should be updated, re-designed, or even the model structure needs to be re-selected.

Up to now, the offline modeling of the distributed predictive modeling framework is finished. All models including the information extraction models inside different blocks and the regression model will be stored in the model library for online utilization. When the new data sample is extracted from the real-time databased, the first step of online key performance index prediction is to obtain the distributed block information from the new data. After the block information has been integrated together, it is imported to the regression model to generate the prediction result for the key performance index. Then, the performance of the prediction result is evaluated. If there is any performance degradation or abnormal event, the main causes need to be identified, which corresponds to the diagnosis step in this framework. Based on the performance diagnosis method, the critical blocks can be identified which takes main responsibility for the abnormality or performance degradation. Furthermore, inside each critical-to-performance block, the contribution of each process variable

Please cite this article in press as: Z. Ge, Distributed predictive modeling framework for prediction and diagnosis of key performance index in plant-wide processes, J. Process Control (2017), http://dx.doi.org/10.1016/j.jprocont.2017.08.010