Chemometrics and Intelligent Laboratory Systems 176 (2018) 89–100

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

An intelligent non-optimality self-recovery method based on reinforcement learning with small data in big data era

Yan Qin [a], Chunhui Zhao [a,*], Furong Gao [b]

[a] State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China
[b] Department of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

A R T I C L E I N F O

Keywords:
Intelligent non-optimality self-recovery
Non-optimal diagnosis
Reinforcement learning
Actor-Critic structure
Big data era

A B S T R A C T

Batch processes have attracted extensive attention as a crucial manufacturing way in modern industries. Although they are well equipped with control devices, batch processes may operate at a non-optimal status because of process disturbances, equipment aging, feedstock variations, etc. As a result, the quality indices or economic benefits may be undesirable using the pre-defined normal operation conditions. And this phenomenon is called non-optimality here. Therefore, it is indispensable to timely remedy the process to its optimal status without accurate models or amounts of data. To solve this problem, this study proposes an intelligent non-optimality self-recovery method based on reinforcement learning. First, the causal variables that lead to the non-optimality are identified by developing a status-degraded Fisher discriminant analysis with consideration of sparsity. Second, on the basis of self-learning mechanism, an intelligent self-recovery method is proposed using the reinforcement learning to automatically adjust the set-points of the causal controlled variables. The self-recovery action is taken iteratively through the Actor-Critic structure with two neural networks. In this way, effective actions are taken to remedy the process to its expected status which only require small data. Finally, the efficacy of the proposed method is illustrated by both numerical case and a typical batch-type manufacturing process, i.e., the injection molding process.

1. Introduction

Batch process, which is an important manufacturing way and characterized by a finite duration and sequential operation steps, has been widely used for producing a number of products in a repetitive manner [1,2]. It has significant advantages to deal with the frequently changing market demand since multiple products can be easily switched without changing manufacturing equipment. This flexible way makes batch processes competitive to produce low-volume and high-value products. Considering that economic benefits are realized by making the batch process operate at the optimal status, timely and reliable operation optimization has played an indispensable role to ensure satisfied operation status.

In fact, batch processes in common suffer from persistent process disturbances and have the inherent time-varying process dynamics, for example, alternation of feedstocks, the absence of equilibrium points [3,4], etc. As a consequence, they are driven from their optimal statuses to a class of non-optimal statuses, in which although all devices work well, the corresponding economic benefits are suboptimal. Here, this phenomenon is termed non-optimality, which is different from process fault that is an unpermitted deviation of the system from the acceptable operation conditions [5]. For continuous processes, this topic has been given much attention in the field of process performance assessment [6–9]. Liu et al. [6] diagnosed abnormal operating conditions using the variable contribution-based method, in which the contribution is ranked by selecting variables having relatively significant influences on the deviation of assessment index that was defined based on each performance grade. However, it is improper to directly apply the existing methods for continuous processes to batch processes regardless of the unique characteristics of batches, such as multiphase [9] and time-varying process [10]. To the best of authors' knowledge, few studies have been reported for batch processes concerning this issue. Besides, for continuous processes, causal variables are diagnosed in the way that is similar to variable contribution analysis in fault diagnosis in previous work [11]. Therefore, it may have the same problem of smearing effect, which may cause misleading results [12]. Moreover, after identifying causal variables, the work mentioned above did not consider whether and how to take effective measures to eliminate non-optimal influences.

* Corresponding author.
E-mail address: chhzhao@zju.edu.cn (C. Zhao).

https://doi.org/10.1016/j.chemolab.2018.03.010
Received 25 January 2018; Received in revised form 9 March 2018; Accepted 20 March 2018
Available online 21 March 2018

For batch processes, the most important objective is to reach the desired product quality at the end of each batch [13]. Therefore, the optimality, i.e., the optimal status, is in common assessed by the batch-end product indices, such as product weight and composition. The repetitive nature of batch processes is usually exploited to update the process trajectories or operation conditions for the current batch by using the batch-end quality error from the previous batch [14]. This is the idea of batch-to-batch control methods. However, the feedback mechanism is implemented until the termination of a batch and it lacks the online quality control ability. Sequentially, more advanced within-batch quality control methods have been developed to reject the within-batch process disturbances and time-varying dynamics [15–18]. Lu et al. [15] predicted the batch-end quality indices at each decision point and the quality control strategy was triggered if quality prediction error was outside the pre-defined region. However, process knowledge is required to describe the relationship between the controlled variables and quality indices to calculate the new set-points of controlled variables. Unfortunately, such priori knowledge may not always be available. In the era of big data, sufficient data make it possible to describe the relationships by developing data-driven process models. Cao et al. [16] proposed a recursive identification method for modeling a batch process by exploiting its intrinsic repeatable pattern. Aumi et al. [17] developed an inferential model for model predictive quality control, in which the final quality was predicted based on a causal model to link the future measurable process variables and state variables. On the basis of this, the control inputs were calculated using a cost function. Corbett et al. [18] identified a state-space model based on subspace identification algorithm to efficiently describe the transient process dynamics. The methods mentioned above require accurate process models to describe the process dynamics, which, however, are static in nature since they are only identified for a specific operating condition. The process dynamics of batch processes may change with time which is frequently influenced by various factors, but the process models cannot be timely updated. It has been pointed out that model mismatch can significantly degrade the performance of the model-based control scheme [19–26]. Considering that developing a new model in a short time is impossible, it brings a significant challenge to timely adjust controlled variables under non-optimal status to ensure satisfied product qualities.

Recently, intelligent manufacturing and industrial 4.0 have attracted extensive attention as a promising way to improve production efficiency and product qualities. Specifically, the industrial processes have recently championed the development of artificial intelligence, in particular, the reinforcement learning (RL) [27]. By observing the behaviors of species interact with their environment, RL was originated from the experimental animal learning and became well known as a method referring to an actor or agent that interacts with its environment and modifies its actions, or control policies, based on reward or punishment received [28]. It is a prevailing way to deal with the decision problem, in particular when the process dynamics is difficult to describe. Besides, it is an online way that small data are required in advance. Therefore, RL has received increasing attention and been widely used in process control field [29–31]. For batch processes, Syafiie et al. [32] successfully controlled batch thermal sterilization of canned food using RL in symbolic state space to track the pre-defined set-points without process models and excellent performance was achieved. Ernst et al. [33] compared the performance of RL with the famous model predictive control (MPC) method in electrical power oscillations damping control problem. Their experiment results showed that RL might be competitive with MPC even in contexts where a good deterministic system model was available. Besides, Frank et al. [28] pointed out that there was a strong connection from a theoretical point of view between adaptive optimal control method and RL.

Motivated by the above analysis, a non-optimality self-recovery method based on RL is proposed for the first time on batch processes. Considering the merits of dealing with dynamic process, RL is suitable to solve the quality control issue of batch processes when they fall into non-optimal statuses. The proposed method is based on the following recognitions: (1) The causal variables that indicate the differences between the non-optimal status and the optimal status can be identified through data analysis; (2) It may remedy the process to its optimal status by adjusting the current set-points of some causal variables; (3) Despite that process dynamics is time-varying and unknown, the set-points of controlled variables can be adjusted based on learning and heuristics. Here, the proposed method concentrates on identifying causal variables that are responsible for the non-optimal status and then the process is brought back to the optimal status based on an online self-recovery adjustment method. First, a sparse status-degraded Fisher discriminant analysis (SSFDA) method is proposed to identify the critical causal variables responsible for the non-optimal status which are termed as causal controlled variables (CCVs). The optimal scoring form is adopted in solving SSFDA by coping with the nonconvex problem caused by sparsity. And the equivalence between these two forms is proved. Second, a non-optimality self-recovery method based on RL is proposed to correct the current set-points of CCVs. It automatically learns what to do and how to map situations to actions to minimize the quantified quality error. It is, in fact, a self-recovery way by calculating a set of system trajectories and instantaneous reward values based on a sequence of learning problems. Here, a heuristic Actor-Critic structure of RL is designed. It includes two neural networks, one of which is used to calculate the set-point of CCVs and the other one is used to evaluate the performance of the concerned set-points. These two neural networks are self-learning online according to the errors between the specified quality indices and the real-time quality information. The contribution of this article is summarized as follows.

(1) It is the first time that a closed-loop framework of non-optimality self-recovery is proposed for batch processes with concurrent consideration of non-optimality diagnosis and operation optimization.
(2) A specific non-optimality diagnosis, i.e., status-degraded FDA, is developed for batch processes.
(3) A non-optimality self-recovery method based on RL is proposed by automatically adjusting set-points of CCVs for batch process recovery to its optimal status.

The organization of the paper is as follows. The details of the proposed method are presented in Section 2, including identification of non-optimal causal variables and a self-recovery strategy based on RL to remedy the process to its optimal status. In Section 3, the feasibility and efficacy of the proposed method are illustrated. Conclusions are drawn in the last section.

2. Methodology

An apparent phenomenon of batch processes is the multiple operation statuses, in which the non-optimal operation caused by improper set-points may lead to unsatisfactory quality indices. In this section, we propose a non-optimality self-recovery method to deal with this problem, which includes two components. First, a non-optimality diagnosis method, termed sparse status-degraded Fisher discriminative analysis (SSFDA) here, is proposed for batch processes at the first time. On the basis of this, causal variables related to the status degradation are directly identified. Second, a non-optimality self-recovery method is proposed on the basis of the self-learning mechanism which can automatically adjust the set-points of the causal controlled variables without the rigorous models or priori knowledge. It can find the optimal operation condition in real-time and thus bring the process to its optimal status.

2.1. Non-optimality diagnosis

To begin with the proposed method, the data structure of batch processes is briefly described. Assuming that batch length is K, a $J_x \times K$